

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:55:41 ; Search time 88 Seconds
(without alignments)
1644.317 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGSDRARKGGGPKDFGAGL.....PGVTRPLPFRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 781663 seqs, 169239676 residues

Total number of hits satisfying chosen parameters: 781663

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4681	100.0	855	6	US-10-612-466B-2
2	4681	100.0	855	7	US-11-104-110-1
3	4681	100.0	855	7	US-11-104-111-22
4	4676	99.9	855	1	PCT-US05-10454-215
5	4676	99.9	855	5	US-09-410-362B-2
6	4676	99.9	855	7	US-11-154-939-650
7	4676	99.9	855	7	US-11-157-575-650
8	4676	99.9	855	8	US-60-687-846-16
9	4676	99.9	855	8	US-60-701-050-901
10	4676	99.9	855	7	US-11-154-939-651
11	4676	99.9	855	7	US-11-167-575-651
12	4676	99.9	855	8	US-60-687-846-15
13	4676	99.9	855	8	US-60-701-050-900
14	4642	99.2	853	6	US-10-530-187-242
15	4319	92.3	932	6	US-10-450-763-51801
16	4175.5	89.2	782	7	US-11-050-926-312
17	1414	30.2	258	7	US-11-104-110-2
18	1414	30.2	258	7	US-11-104-111-23
19	1352	28.9	362	6	US-10-450-763-51800
20	1319	28.2	241	5	US-09-410-362C-2
21	1319	28.2	241	5	US-09-410-362B-3
22	1319	28.2	241	5	US-09-410-362B-35
23	1319	28.2	241	6	US-10-612-466B-4
24	1124	24.0	811	8	US-60-685-372-1264
25	1124	24.0	811	8	US-60-692-282-314

26	1122	24.0	802	5	US-09-978-191C-169	Sequence 169, App
27	1122	24.0	802	7	US-11-129-762-169	Sequence 169, App
28	1122	24.0	802	7	US-11-037-243-113	Sequence 113, App
29	1113.5	23.8	850	7	US-11-037-243-108	Sequence 108, App
30	990.5	21.2	453	6	US-10-450-763-51797	Sequence 51797, A
31	890.5	19.0	152	5	US-09-410-362C-8	Sequence 8, Appli
32	890.5	19.0	152	5	US-09-410-362B-41	Sequence 41, Appl
33	816.5	17.4	152	5	US-09-410-362C-10	Sequence 10, Appl
34	816.5	17.4	152	5	US-09-410-362B-43	Sequence 43, Appl
35	718.5	15.3	1128	7	US-11-037-243-97	Sequence 97, Appl
36	717.5	15.3	531	6	US-10-522-668-2	Sequence 2, Appli
37	692	14.8	1042	1	PCT-US05-06052-1	Sequence 1, Appli
38	692	14.8	1042	6	US-10-543-003-4	Sequence 4, Appli
39	692	14.8	1042	7	US-11-067-811-1	Sequence 1, Appli
40	682.5	14.6	1019	8	US-60-710-726-1836	Sequence 1836, Ap
41	679.5	14.5	1019	1	PCT-US03-10870-2243	Sequence 2243, Ap
42	677	14.5	126	5	US-09-410-362B-22	Sequence 22, Appl
43	663.5	14.2	1113	1	PCT-US05-06052-4	Sequence 4, Appli
44	663.5	14.2	1113	6	US-10-543-003-3	Sequence 3, Appli
45	663.5	14.2	1113	7	US-11-067-811-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-612-466B-2
; Sequence 2, Application US/10612466B
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; APPLICANT: Yeh, Juinn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 16, THE
; FILE REFERENCE: 24745-1625
; CURRENT APPLICATION NUMBER: US/10/612,466B
; CURRENT FILING DATE: 2003-07-01
; PRIOR APPLICATION NUMBER: 60/394,347
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-612-466B-2

Query Match	100.0%;	Score	4681;	DB	6;	Length	855;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	855;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
QY	1	MGS	DRARKGGGPKDFGAGLKYNSRHEKYNGLGEEVFLPVNNVKVKGHPGRVWVLA	60			
DB	1	MGS	DRARKGGGPKDFGAGLKYNSRHEKYNGLGEEVFLPVNNVKVKGHPGRVWVLA	60			
QY	61	VLIGLLLVLLGIGFLVHLYQYDRVQKVFNGYMRITNENFVDAYENSSTFVSLASKV	120				
DB	61	VLIGLLLVLLGIGFLVHLYQYDRVQKVFNGYMRITNENFVDAYENSSTFVSLASKV	120				
QY	121	KDALKLLISGVFPGLPYHKESAVTAFSEGSVAYTWSEFSIQHLVEEAERVMAERVVM	180				
DB	121	KDALKLLISGVFPGLPYHKESAVTAFSEGSVAYTWSEFSIQHLVEEAERVMAERVVM	180				
QY	181	LPPRARSLKSFVVTWSVAFPTDSKTQRTQDNSCSFGLHARGVELMRFTTGPDPSPYPA	240				
DB	181	LPPRARSLKSFVVTWSVAFPTDSKTQRTQDNSCSFGLHARGVELMRFTTGPDPSPYPA	240				
QY	241	HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVVTNTLSPMPEHALVOLCGTYPPS	300				
DB	241	HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVVTNTLSPMPEHALVOLCGTYPPS	300				
QY	301	YNLTHSSQNVLLITLITNTERRHPGFETFFOLPMSSCGGRLRKAQQTNSPYPGHY	360				

Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGRLRKAQGTNSPYYPGHY 360
Qy 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Db 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Qy 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Qy 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
Db 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
Qy 601 DCDCGLRSFTRQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Qy 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTTCEML 780
Db 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTTCEML 780
Qy 781 PQOITPRMCMVGLSGVSDSCQSDGSLSSVEADGRIFQAGVSWGDCQAQRNKPQYIT 840
Db 781 PQOITPRMCMVGLSGVSDSCQSDGSLSSVEADGRIFQAGVSWGDCQAQRNKPQYIT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 2

US-11-104-110-1
; Sequence 1, Application US/11104110
; GENERAL INFORMATION:
; APPLICANT: Ruggles, Sandra
; APPLICANT: Nguyen, Jack
; TITLE OF INVENTION: CLEAVAGE OF VEGF AND VEGF RECEPTOR BY WILDTYPE AND MUTANT MT-SPI
; FILE REFERENCE: 25840-502
; CURRENT APPLICATION NUMBER: US/11/104,110
; PRIOR FILING DATE: 2005-04-12
; PRIOR APPLICATION NUMBER: 60/561,720
; PRIOR FILING DATE: 2004-04-12
; PRIOR APPLICATION NUMBER: 10/677,977
; PRIOR FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,388
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-104-110-1

Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGSRRARKGCGGKDFGAGLKYNSRHEKVGLEGEVEFLPNNVKKVEKHGPGRWVLA 60
Db 1 MGSRRARKGCGGKDFGAGLKYNSRHEKVGLEGEVEFLPNNVKKVEKHGPGRWVLA 60
Qy 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKFNMGYMRITNENFVDAYENSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKFNMGYMRITNENFVDAYENSNSTEFVSLASKV 120
Qy 121 KDALKLIYSYVPPFLGPHKESAVTAFSEGSAVIATYNSSEFSIPOHLVEAEAVMAEERVYM 180
Db 121 KDALKLIYSYVPPFLGPHKESAVTAFSEGSAVIATYNSSEFSIPOHLVEAEAVMAEERVYM 180
Qy 181 LPPRARSLSKSVTSVVAFFPTDSKTIVORTDONSQSGFLHARGVELMRFTTTPGFPDPSYPA 240
Db 181 LPPRARSLSKSVTSVVAFFPTDSKTIVORTDONSQSGFLHARGVELMRFTTTPGFPDPSYPA 240
Qy 241 HARQWALRGDADSVLSLTPRSFDLASCDERGSDLVTYNTLSMPHEHALVOLCGTYPPS 300
Db 241 HARQWALRGDADSVLSLTPRSFDLASCDERGSDLVTYNTLSMPHEHALVOLCGTYPPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGRLRKAQGTNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGRLRKAQGTNSPYYPGHY 360
Qy 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Db 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Qy 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYSSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Qy 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
Db 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
Qy 601 DCDCGLRSFTRQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Qy 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTTCEML 780
Db 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTTCEML 780
Qy 781 PQOITPRMCMVGLSGVSDSCQSDGSLSSVEADGRIFQAGVSWGDCQAQRNKPQYIT 840
Db 781 PQOITPRMCMVGLSGVSDSCQSDGSLSSVEADGRIFQAGVSWGDCQAQRNKPQYIT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 3

US-11-104-111-22
; Sequence 22, Application US/11104111
; GENERAL INFORMATION:
; APPLICANT: Waugh Ruggles, Sandra
; APPLICANT: Nguyen, Jack
; TITLE OF INVENTION: CLEAVAGE OF VEGF AND VEGF RECEPTOR BY WILD-TYPE AND MUTANT
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: 25840-503
; CURRENT APPLICATION NUMBER: US/11/104,111
; CURRENT FILING DATE: 2005-04-12
; PRIOR APPLICATION NUMBER: 60/561,671
; PRIOR FILING DATE: 2004-04-12
; PRIOR APPLICATION NUMBER: 10/677,977
; PRIOR FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,388
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 22
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-104-111-22

```
Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDPFGAGLKYNRHEKVGLEEGVEFLPNNVKKVEKHGGRWVLA 60
DB 1 MGSDRARGGGGKDPFGAGLKYNRHEKVGLEEGVEFLPNNVKKVEKHGGRWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
DB 61 VLIGLLLVLLGIGFLVHLYQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPOHLVBEAERVMAERVVM 180
DB 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPOHLVBEAERVMAERVVM 180
QY 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
DB 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTYPPS 300
DB 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
DB 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKVFYLLPEPGVAGTCTPKDYVEINGEKYCERSQFVVTS 420
DB 361 PPNIDCTWNIIEVNNQHVKVFYLLPEPGVAGTCTPKDYVEINGEKYCERSQFVVTS 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
QY 481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDGDNDSDEQSCPAQTFPCSNKGKCLSK 540
DB 481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDGDNDSDEQSCPAQTFPCSNKGKCLSK 540
QY 541 SOQCNGKDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
DB 541 SOQCNGKDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
QY 601 DDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQERRLKRILISHPFNFDFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQERRLKRILISHPFNFDFDYDIALLELEKP 720
QY 721 AYSMSWRPILCPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780
DB 721 AYSMSWRPILCPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780
QY 781 PQOITPRMCMVGLSGGVDSCQSGGGLSSVEADGRIFQAGVVSWSGCAQRNKPQYIT 840
DB 781 PQOITPRMCMVGLSGGVDSCQSGGGLSSVEADGRIFQAGVVSWSGCAQRNKPQYIT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855
```

RESULT 4

PCT-US05-10454-215
; Sequence 215, Application PC/TUS0510454
; GENERAL INFORMATION:
; APPLICANT: Ford, Shirin K.
; APPLICANT: Perkins, Nancy-Anne A.
; APPLICANT: Jackson, Donald G.
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO EPIDERMAL
; FILE OF INVENTION: GROWTH FACTOR RECEPTOR MODULATORS IN NON-SMALL CELL LUNG CANCER
; FILE REFERENCE: 10219 PCT
; CURRENT APPLICATION NUMBER: PCT/US05/10454
; CURRENT FILING DATE: 2005-04-05
; PRIOR APPLICATION NUMBER: 60/556,903
; PRIOR FILING DATE: 2004-03-26
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 215
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US05-10454-215

```
Query Match 99.9%; Score 4676; DB 1; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDPFGAGLKYNRHEKVGLEEGVEFLPNNVKKVEKHGGRWVLA 60
DB 1 MGSDRARGGGGKDPFGAGLKYNRHEKVGLEEGVEFLPNNVKKVEKHGGRWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
DB 61 VLIGLLLVLLGIGFLVHLYQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPOHLVBEAERVMAERVVM 180
DB 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPOHLVBEAERVMAERVVM 180
QY 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
DB 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTYPPS 300
DB 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
DB 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKVFYLLPEPGVAGTCTPKDYVEINGEKYCERSQFVVTS 420
DB 361 PPNIDCTWNIIEVNNQHVKVFYLLPEPGVAGTCTPKDYVEINGEKYCERSQFVVTS 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
QY 481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDGDNDSDEQSCPAQTFPCSNKGKCLSK 540
DB 481 SDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDGDNDSDEQSCPAQTFPCSNKGKCLSK 540
QY 541 SOQCNGKDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
DB 541 SOQCNGKDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
QY 601 DDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQERRLKRILISHPFNFDFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQERRLKRILISHPFNFDFDYDIALLELEKP 720
```

```
QY 721 AYSVMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCEMLL 780
Db 721 AYSVMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCEMLL 780
QY 781 PQOITPRMCMVGLSGVDSQCGDGLSSVEADGRIFQAGVSWGDCGAQRNKPQYVT 840
Db 781 PQOITPRMCMVGLSGVDSQCGDGLSSVEADGRIFQAGVSWGDCGAQRNKPQYVT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 5
US-09-410-362E-2
; Sequence 2, Application US/09410362E
; GENERAL INFORMATION:
; APPLICANT: CRAIK, CHARLES S.
; APPLICANT: YAKUCHI, TOSHIHIKO
; APPLICANT: SCHUMAN, MARC
; TITLE OF INVENTION: MEMBRANE TYPE SERINE PROTEASE 1 (MT-SPI) AND USES THEREOF
; FILE REFERENCE: 28644-701.201
; CURRENT APPLICATION NUMBER: US/09/410,362E
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-410-362E-2

Query Match 99.9%; Score 4676; DB 5; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGGPKDFGAGLKYNRHEKVNGLKEGVEFLPVNNVKKVKEKGGPRWVLA 60
Db 1 MGSDRARKGGGGPKDFGAGLKYNRHEKVNGLKEGVEFLPVNNVKKVKEKGGPRWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFDAYENSNSFEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFDAYENSNSFEFVSLASKV 120
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
QY 181 LPPRARSLSKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYPA 240
Db 181 LPPRARSLSKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLTVVNTLSMPHEHALVOLCGTYP 300
Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLTVVNTLSMPHEHALVOLCGTYP 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
Db 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
QY 421 NSNKITVRFHSDQSYDTGFLAEVLSYSDSPCGOFTCRGRCIRKELRCGWDGADCTDH 480
Db 421 NSNKITVRFHSDQSYDTGFLAEVLSYSDSPCGOFTCRGRCIRKELRCGWDGADCTDH 480
QY 481 SDELNCSADAGHQFTCKNFKCPFLFWCDNVNDGDNDEQGCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSADAGHQFTCKNFKCPFLFWCDNVNDGDNDEQGCSPAQTFRCNSGKCLSK 540
QY 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDGSD 600
```

```
Db 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDGSD 600
QY 601 DCDGLRSFTRQARVVGTTDADEGEWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
Db 601 DCDGLRSFTRQARVVGTTDADEGEWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDOSQSAFGVQERRLRKRIISHPFFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDOSQSAFGVQERRLRKRIISHPFFNDFTDYDIALLELEKP 720
QY 721 AYSVMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCEMLL 780
Db 721 AYSVMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCEMLL 780
QY 781 PQOITPRMCMVGLSGVDSQCGDGLSSVEADGRIFQAGVSWGDCGAQRNKPQYVT 840
Db 781 PQOITPRMCMVGLSGVDSQCGDGLSSVEADGRIFQAGVSWGDCGAQRNKPQYVT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 6
US-11-154-939-650
; Sequence 650, Application US/11154939
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/154,939
; CURRENT FILING DATE: 2005-06-17
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 650
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-154-939-650

Query Match 99.9%; Score 4676; DB 7; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGGPKDFGAGLKYNRHEKVNGLKEGVEFLPVNNVKKVKEKGGPRWVLA 60
Db 1 MGSDRARKGGGGPKDFGAGLKYNRHEKVNGLKEGVEFLPVNNVKKVKEKGGPRWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFDAYENSNSFEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFDAYENSNSFEFVSLASKV 120
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
QY 181 LPPRARSLSKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYPA 240
Db 181 LPPRARSLSKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLTVVNTLSMPHEHALVOLCGTYP 300
Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLTVVNTLSMPHEHALVOLCGTYP 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
Db 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
```

```
QY 421 NSNKITVRHSDQSYTDGFLAEYLSYSSDDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
Db |||||
QY 421 NSNKITVRHSDQSYTDGFLAEYLSYSSDDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
Db |||||
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db |||||
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db |||||
QY 541 SQQNGKDDCGDSDSEASCPKVVVCTKHTYRCLNGCLSLKGNPECDGKEDCDSDGDEK 600
Db |||||
QY 541 SQQNGKDDCGDSDSEASCPKVVVCTKHTYRCLNGCLSLKGNPECDGKEDCDSDGDEK 600
Db |||||
QY 601 DCDGLRSFTROARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db |||||
QY 601 DCDGLRSFTROARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db |||||
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db |||||
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db |||||
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db |||||
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db |||||
QY 781 PQITPRMVCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db |||||
QY 781 PQITPRMVCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db |||||
QY 841 RLPLFRDWIKENTGV 855
Db |||||
QY 841 RLPLFRDWIKENTGV 855
Db |||||

RESULT 7
US-11-167-575-650
; Sequence 650, Application US/11167575
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/167,575
; CURRENT FILING DATE: 2005-06-28
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 650
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-167-575-650

Query Match 99.9%; Score 4676; DB 7; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGKDFGAGLKYNRSHKXNGLEGEVFLPVNNVKKVKGHPGRWVLA 60
Db |||||
QY 1 MGSDRARKGGGKDFGAGLKYNRSHKXNGLEGEVFLPVNNVKKVKGHPGRWVLA 60
Db |||||
QY 61 VLIGLLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db |||||
QY 61 VLIGLLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db |||||
QY 121 KDALKLLYSGVPFLGPHYHKS AVTAFSEGSVIAYWSEFSIPQHLVEAEERVMABEERVVM 180
Db |||||
QY 121 KDALKLLYSGVPFLGPHYHKS AVTAFSEGSVIAYWSEFSIPQHLVEAEERVMABEERVVM 180
Db |||||
QY 181 LPPRARSLSFVVTSVVAFPTDSKTQVTDQNSCSFGLHARGVELMFTTPGPPDSPYPA 240
Db |||||
QY 181 LPPRARSLSFVVTSVVAFPTDSKTQVTDQNSCSFGLHARGVELMFTTPGPPDSPYPA 240
Db |||||
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDEGSDLVTVNTLSPMEPHALVQLCGTYPPS 300
Db |||||
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDEGSDLVTVNTLSPMEPHALVQLCGTYPPS 300
Db |||||
```

```
QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFFOLPRMSSCGGLRKAQGTFSNPYPGHY 360
Db |||||
QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFFOLPRMSSCGGLRKAQGTFSNPYPGHY 360
Db |||||
QY 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLRPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
Db |||||
QY 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLRPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
Db |||||
QY 421 NSNKITVRHSDQSYTDGFLAEYLSYSSDDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
Db |||||
QY 421 NSNKITVRHSDQSYTDGFLAEYLSYSSDDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
Db |||||
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db |||||
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db |||||
QY 541 SQQNGKDDCGDSDSEASCPKVVVCTKHTYRCLNGCLSLKGNPECDGKEDCDSDGDEK 600
Db |||||
QY 541 SQQNGKDDCGDSDSEASCPKVVVCTKHTYRCLNGCLSLKGNPECDGKEDCDSDGDEK 600
Db |||||
QY 601 DCDGLRSFTROARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db |||||
QY 601 DCDGLRSFTROARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db |||||
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db |||||
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db |||||
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db |||||
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db |||||
QY 781 PQITPRMVCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db |||||
QY 781 PQITPRMVCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db |||||
QY 841 RLPLFRDWIKENTGV 855
Db |||||
QY 841 RLPLFRDWIKENTGV 855
Db |||||

RESULT 8
US-60-687-846-16
; Sequence 16, Application US/60687846
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001606
; CURRENT APPLICATION NUMBER: US/60/687,846
; CURRENT FILING DATE: 2005-06-07
; NUMBER OF SEQ ID NOS: 1565
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-687-846-16

Query Match 99.9%; Score 4676; DB 8; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGKDFGAGLKYNRSHKXNGLEGEVFLPVNNVKKVKGHPGRWVLA 60
Db |||||
QY 1 MGSDRARKGGGKDFGAGLKYNRSHKXNGLEGEVFLPVNNVKKVKGHPGRWVLA 60
Db |||||
QY 61 VLIGLLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db |||||
QY 61 VLIGLLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db |||||
QY 121 KDALKLLYSGVPFLGPHYHKS AVTAFSEGSVIAYWSEFSIPQHLVEAEERVMABEERVVM 180
Db |||||
```

```

Db      121 KDALKLLYSVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMAERVVM 180
Qy      181 LPPRARSLSKSVVTSVVAFTDSTKVTQRTQDNSCSFGLHARGVELMRFTTTPGFPDPSYPYA 240
Db      181 LPPRARSLSKSVVTSVVAFTDSTKVTQRTQDNSCSFGLHARGVELMRFTTTPGFPDPSYPYA 240
Qy      241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
Db      241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
Qy      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
Qy      361 PPNIDCTWNIIEVNNQHVSKFVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db      361 PPNIDCTWNIIEVNNQHVSKFVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy      421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db      421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Qy      481 SDELNCSCDAGHOFCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQOTFRCSNGKCLSK 540
Db      481 SDELNCSCDAGHOFCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQOTFRCSNGKCLSK 540
Qy      541 SOQNGKDDCGSDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Db      541 SOQNGKDDCGSDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Qy      601 DDCGLRSFTRQARVVGTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db      601 DDCGLRSFTRQARVVGTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy      661 DRGFYSDPTQWTAFLGLHDQSORAPGVQBRRLKRIISHPFENDFTFDYDIALLELEKP 720
Db      661 DRGFYSDPTQWTAFLGLHDQSORAPGVQBRRLKRIISHPFENDFTFDYDIALLELEKP 720
Qy      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

```

RESULT 9

```

US-60-701-050-901
; Sequence 901, Application US/60701050
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CLO01621
; CURRENT APPLICATION NUMBER: US/60/701,050
; CURRENT FILING DATE: 2005-07-21
; NUMBER OF SEQ ID NOS: 2554
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 901
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-701-050-901

```

```

Query Match      99.9%; Score 4676; DB 8; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLSEGVFLPVNNVKVKEKHGPRWVLA 60

```

```

Db      1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLSEGVFLPVNNVKVKEKHGPRWVLA 60
Qy      61 VLIGLLVLLGIGLGVHQLQYRDVQKVQFNGYWRITNENFVDAYENSNSTEFVSLASKV 120
Db      61 VLIGLLVLLGIGLGVHQLQYRDVQKVQFNGYWRITNENFVDAYENSNSTEFVSLASKV 120
Qy      121 KDALKLLYSVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMAERVVM 180
Db      121 KDALKLLYSVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMAERVVM 180
Qy      181 LPPRARSLSKSVVTSVVAFTDSTKVTQRTQDNSCSFGLHARGVELMRFTTTPGFPDPSYPYA 240
Db      181 LPPRARSLSKSVVTSVVAFTDSTKVTQRTQDNSCSFGLHARGVELMRFTTTPGFPDPSYPYA 240
Qy      241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
Db      241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
Qy      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
Qy      361 PPNIDCTWNIIEVNNQHVSKFVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db      361 PPNIDCTWNIIEVNNQHVSKFVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy      421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db      421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Qy      481 SDELNCSCDAGHOFCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQOTFRCSNGKCLSK 540
Db      481 SDELNCSCDAGHOFCKNFKCPLEFWVCDVNDGDNDSDEQSCCPAQOTFRCSNGKCLSK 540
Qy      541 SOQNGKDDCGSDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Db      541 SOQNGKDDCGSDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Qy      601 DDCGLRSFTRQARVVGTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db      601 DDCGLRSFTRQARVVGTDADGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy      661 DRGFYSDPTQWTAFLGLHDQSORAPGVQBRRLKRIISHPFENDFTFDYDIALLELEKP 720
Db      661 DRGFYSDPTQWTAFLGLHDQSORAPGVQBRRLKRIISHPFENDFTFDYDIALLELEKP 720
Qy      721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCCNLL 780
Db      721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCCNLL 780
Qy      781 PQOITPRMCMCVGLSGVDSCQDGGPLSSVEADGRIFOAGVVSWDGCAQRNKPQVYT 840
Db      781 PQOITPRMCMCVGLSGVDSCQDGGPLSSVEADGRIFOAGVVSWDGCAQRNKPQVYT 840
Qy      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

```

RESULT 10

```

US-11-154-939-651
; Sequence 651, Application US/11154939
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CLO01529
; CURRENT APPLICATION NUMBER: US/11/154,939
; CURRENT FILING DATE: 2005-06-17
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 651
; LENGTH: 856

```

	TYPE: PR	ORGANISM: Homo sapiens	US-11-154-939-651
Query Match	99.9%	Score 4676; DB 7; Length 856;	
Best Local Similarity	99.9%	Pred. No. 0;	
Matches	854; Conservative	0; Mismatches	1; Indels
		0; Gaps	0;
Qy	1	MGSDRARKGGGGPKDFGAGLKYNSRHEKVGLEGGVEFLPVNNVKYKEKGGPGRWVLA	60
Db	1	MGSDRARKGGGGPKDFGAGLKYNSRHEKVGLEGGVEFLPVNNVKYKEKGGPGRWVLA	60
Qy	61	VLIGLLVLLIGLGFVWHQLYQDVRVQKVFNGVMRIITNENFVDAYENSNSTEFVSLASKV	120
Db	61	VLIGLLVLLIGLGFVWHQLYQDVRVQKVFNGVMRIITNENFVDAYENSNSTEFVSLASKV	120
Qy	121	KDALKLLYSGVPFLGPHYKESAVTAFSEGSVIATYVSEFSIPOHLVBEAERVMAEERVVM	180
Db	121	KDALKLLYSGVPFLGPHYKESAVTAFSEGSVIATYVSEFSIPOHLVBEAERVMAEERVVM	180
Qy	181	LPPARSLSKSVVTSVVAFFPDSKTQVORTQDSCSFGLHARGVELMFTTTPGPPDSPYP	240
Db	181	LPPARSLSKSVVTSVVAFFPDSKTQVORTQDSCSFGLHARGVELMFTTTPGPPDSPYP	240
Qy	241	HARQWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSPEMPHALVOLCGTYP	300
Db	241	HARQWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSPEMPHALVOLCGTYP	300
Qy	301	YNLTFHSSQNVLITLINTERRHPGFEATFQLPRMSSCGRLRKACQGTNPSYPYGHY	360
Db	301	YNLTFHSSQNVLITLINTERRHPGFEATFQLPRMSSCGRLRKACQGTNPSYPYGHY	360
Qy	361	PPNIDCTWNIENVNNQHKVSKPKPYLLEPGVPAGTCKDYVEINGEKYCGERSQFV	420
Db	361	PPNIDCTWNIENVNNQHKVSKPKPYLLEPGVPAGTCKDYVEINGEKYCGERSQFV	420
Qy	421	NSNKITVRFHSDQSYTDTGLAEYLSYDSSDCPGQFTCTGRCIRKELRCDGWADCTD	480
Db	421	NSNKITVRFHSDQSYTDTGLAEYLSYDSSDCPGQFTCTGRCIRKELRCDGWADCTD	480
Qy	481	SDELNCSDAGHQFTCKNKFCCKPLFWVCDSDVNDCGDSDSQSCPAQTFRCNSGKCL	540
Db	481	SDELNCSDAGHQFTCKNKFCCKPLFWVCDSDVNDCGDSDSQSCPAQTFRCNSGKCL	540
Qy	541	SOQCNKGKDDCGDSDSEASCPCVNVVCTKHTYRCNLGLCLSKGNPECDGKDCSDG	600
Db	541	SOQCNKGKDDCGDSDSEASCPCVNVVCTKHTYRCNLGLCLSKGNPECDGKDCSDG	600
Qy	601	DCDCGLRSFTQARVWGTTDADEGEWPQVSLHALGQGHICGASLISPNMLVSAAHCY	660
Db	601	DCDCGLRSFTQARVWGTTDADEGEWPQVSLHALGQGHICGASLISPNMLVSAAHCY	660
Qy	661	DRGPRYSDPTQWTAFLGHDSQSRSASGVQERRLKRISHPFFNDFDYDIALLELEK	720
Db	661	DRGPRYSDPTQWTAFLGHDSQSRSASGVQERRLKRISHPFFNDFDYDIALLELEK	720
Qy	721	AEYSSVMRPICLPDASHVFPAGKAIWTTGHTQYGGTGALILQKGIIRVINOTTEN	780
Db	721	AEYSSVMRPICLPDASHVFPAGKAIWTTGHTQYGGTGALILQKGIIRVINOTTEN	780
Qy	781	PQQTTPRMCMCVGLISGGVDSQCGDSGGLSSVEADGRIFOAGVVSWSGDCAQRNKP	840
Db	781	PQQTTPRMCMCVGLISGGVDSQCGDSGGLSSVEADGRIFOAGVVSWSGDCAQRNKP	840
Qy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855

```

; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/167,575
; CURRENT FILING DATE: 2005-06-28
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 651
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-167-575-651

```

RESULT 11
US-11-167-575-651
; Sequence 651, Application US/11167575
; GENERAL INFORMATION:


```
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 12
US-60-687-846-15
; Sequence 15, Application US/60687846
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001606
; CURRENT APPLICATION NUMBER: US/60/687,846
; CURRENT FILING DATE: 2005-06-07
; NUMBER OF SEQ ID NOS: 1565
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-687-846-15

Query Match 99.9%; Score 4676; DB 8; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVKGHGPGRWVLLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVKGHGPGRWVLLAA 60

QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KQALKLYSGVPLGYPYHKS AVTA FSEGSVIAYYWFSEFSIPQHLVEAEARVMAERVVM 180
Db 121 KQALKLYSGVPLGYPYHKS AVTA FSEGSVIAYYWFSEFSIPQHLVEAEARVMAERVVM 180

QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGS DLVTYNTLSMPHEPHALVOLCGTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGS DLVTYNTLSMPHEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLTITNERRHPGFEATFFQLPMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLTITNERRHPGFEATFFQLPMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNI DCTWNI EVPNNOHVKSFKFYLLPEPGVPAGTCPKDYVEINGEKYCGERSQFVVTs 420
Db 361 PPNI DCTWNI EVPNNOHVKSFKFYLLPEPGVPAGTCPKDYVEINGEKYCGERSQFVVTs 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SQQCNKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 541 SQQCNKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600

QY 601 DCDCLGRSFTQARVVGGTADGEPWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCLGRSFTQARVVGGTADGEPWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRIISHPPFNDFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRIISHPPFNDFDYDIALLELEKP 720
```

```
QY 721 AESSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCEULL 780
Db 721 AESSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCEULL 780

QY 781 PQQITPRMVCVGLSGGVDSCQSGGGLSSVEADGRIFQAGVWSWGDCGAQRNPGVYT 840
Db 781 PQQITPRMVCVGLSGGVDSCQSGGGLSSVEADGRIFQAGVWSWGDCGAQRNPGVYT 840

QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 13
US-60-701-050-900
; Sequence 900, Application US/60701050
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001621
; CURRENT APPLICATION NUMBER: US/60/701,050
; CURRENT FILING DATE: 2005-07-21
; NUMBER OF SEQ ID NOS: 2554
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 900
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-701-050-900

Query Match 99.9%; Score 4676; DB 8; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVKGHGPGRWVLLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVKGHGPGRWVLLAA 60

QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KQALKLYSGVPLGYPYHKS AVTA FSEGSVIAYYWFSEFSIPQHLVEAEARVMAERVVM 180
Db 121 KQALKLYSGVPLGYPYHKS AVTA FSEGSVIAYYWFSEFSIPQHLVEAEARVMAERVVM 180

QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGS DLVTYNTLSMPHEPHALVOLCGTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGS DLVTYNTLSMPHEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLTITNERRHPGFEATFFQLPMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLTITNERRHPGFEATFFQLPMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNI DCTWNI EVPNNOHVKSFKFYLLPEPGVPAGTCPKDYVEINGEKYCGERSQFVVTs 420
Db 361 PPNI DCTWNI EVPNNOHVKSFKFYLLPEPGVPAGTCPKDYVEINGEKYCGERSQFVVTs 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SQQCNKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 541 SQQCNKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
```


Db 541 SQCCNGKDCGDSDEASCPKVNVTCTKHYRCLNGLCLSKGNPECDKEDCSGDSBK 600
Qy 601 DDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYD 660
Db 601 DDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYD 660
Qy 661 DRGFRYSDPTQWTAFLGLHDQSQSAPGVOERRLKRIISHPFFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSQSAPGVOERRLKRIISHPFFNDFTDYDIALLELEKP 720
Qy 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTICALIQKEIRVINQTTCEMLL 780
Db 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTICALIQKEIRVINQTTCEMLL 780
Qy 781 PQQITPRMCMVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Db 781 PQQITPRMCMVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 14
US-10-530-187-242
; Sequence 242, Application US/10530187
; GENERAL INFORMATION:
; APPLICANT: THE, Bin Tean
; APPLICANT: TAKAHASHI, Masayuki
; TITLE OF INVENTION: Molecular Subclassification of Kidney Tumors and the Discovery of
; FILE REFERENCE: 28927-0014
; CURRENT APPLICATION NUMBER: US/10/530,187
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US2003/031476
; PRIOR FILING DATE: 2003-10-06
; PRIOR APPLICATION NUMBER: US 60/415,775
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 296
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 242
; LENGTH: 853
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-530-187-242

Query Match 99.2%; Score 4642; DB 6; Length 853;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
Qy 1 MGSDBARKGGGKDPFAGLKYNSRHEKVGLEEGVEFLPVNNVKKVEKHGPGRWVLLA 60
Db 1 MGSDBARKGGGKDPFAGLKYNSRHEKVGLEEGVEFLPVNNVKKVEKHGPGRWVLLA 60
Qy 61 VLIGLLLVLLGTLFWLHQLQVDRVQVFNQYMRITNENFDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLLVLLGTLFWLHQLQVDRVQVFNQYMRITNENFDAYENSNSSTEFVSLASKV 120
Qy 121 KDALKLLYSGVPLFPGYHKESAVTAFSGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
Db 121 KDALKLLYSGVPLFPGYHKESAVTAFSGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
Qy 181 LPPRARSLSKSVTSVVAFFTDSKTQVQTDNSCSFGLHARGVELMRTTTFGPDSPYPA 240
Db 181 LPPRARSLSKSVTSVVAFFTDSKTQVQTDNSCSFGLHARGVELMRTTTFGPDSPYPA 239
Qy 241 HARCOWALRGDADSVLSTFTFSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTTPPS 300
Db 240 HA-COWALRGDADSVLSTFTFSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTTPPS 298
Qy 301 YNLTFHSSONVLLITLTINTERRHPGFATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
Db 299 YNLTFHSSONVLLITLTINTERRHPGFATFFQLPRMSSCGRLRKAQGTNSPYPGHY 358

Qy 361 PPNIDCTWNIIEVPNNQHVVKFYLELPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
Db 359 PPNIDCTWNIIEVPNNQHVVKFYLELPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 418
Qy 421 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPCPQFTCRTRGRCIRKELRCGWDCTDH 480
Db 419 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPCPQFTCRTRGRCIRKELRCGWDCTDH 478
Qy 481 SDELNCSDAGHQTCKNFKCKPLFWVCDSDVNDGDNDSDEQSCPAQTRFCSNGKCLSK 540
Db 479 SDELNCSDAGHQTCKNFKCKPLFWVCDSDVNDGDNDSDEQSCPAQTRFCSNGKCLSK 538
Qy 541 SQCCNGKDDCGDSDEASCPKVNVTCTKHYRCLNGLCLSKGNPECDKEDCSGDSBK 600
Db 539 SQCCNGKDDCGDSDEASCPKVNVTCTKHYRCLNGLCLSKGNPECDKEDCSGDSBK 598
Qy 601 DDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYD 660
Db 599 DDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYD 658
Qy 661 DRGFRYSDPTQWTAFLGLHDQSQSAPGVOERRLKRIISHPFFNDFTDYDIALLELEKP 720
Db 659 DRGFRYSDPTQWTAFLGLHDQSQSAPGVOERRLKRIISHPFFNDFTDYDIALLELEKP 718
Qy 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTICALIQKEIRVINQTTCEMLL 780
Db 719 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTICALIQKEIRVINQTTCEMLL 778
Qy 781 PQQITPRMCMVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Db 779 PQQITPRMCMVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 838
Qy 841 RLPLFRDWIKENTGV 855
Db 839 RLPLFRDWIKENTGV 853

RESULT 15
US-10-450-763-51801
; Sequence 51801, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 51801
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (715)..(733)
; OTHER INFORMATION: Kringle domain proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00021B, p-value=5.696e-17, raw score of 13.3
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(932)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-51801

Query Match 92.3%; Score 4319; DB 6; Length 932;
Best Local Similarity 96.0%; Pred. No. 2.2e-313;
Matches 800; Conservative 7; Mismatches 22; Indels 4; Gaps 4;

QY 27 EKNGLEEGVEFLPVNNVKKVKGGRWVLAALVIGLLLVLLGIGFLVWHLQYRDVRV 86
 Db :|||||
 100 QKNGLEEGVEFLPVNNVKKVKGGRWVLAALVIGLLLVLLGIGFLVWHLQYRDVRV 159
 QY 87 QKVFNGYMRITNENFVDAVENSENSTEFVSLASKVDALKLLYSGVFFLGPYHKESAVTAF 146
 Db :|||||
 160 QKVFNGYMRITNENFVDAVENSENSTEFVSLASKVDALKLLYSGVFFLGPYHKESAVTAF 219
 QY 147 SEGSIAYYWSFSLPOHLVEEAERVMAEERVVMPLPPRARSLSKFVTSVVAFPDTSKTIV 206
 Db :|||||
 220 SEGSIAYYWSFSLPOHLVEEAERVMAEERVVMPLPPRARSLSKFVTSVVAFPDTSKTIV 279
 QY 207 QRTQNSCSFGLHARGVELMRTFTTFCPPDSPYPAHARQWALRGDADSVLSLTFRSFDLA 266
 Db :|||||
 280 QRTQNSCSFGLHARGVELMRTFTTFCPPDSPYPAHARQWALRGDADSVLSLTFRSFDLA 339
 QY 267 SCDEGRSLVTVYNTLSMPHEHALVOLCGTYPPSYNLTFHSSQNVLLITLINTERRHPG 326
 Db :|||||
 340 SCDEGRSLVTVYNTLSMPHEHALVOLCGTYPPSYNLTFHSSQNVLLITLINTERRHPG 399
 QY 327 PEATFFQLPRMSSCGRLKKAQGTNSPYPGHYPP-NIDCTWNIENPNNOHVKVSFKFF 385
 Db :|||||
 400 FEATFFQLPRMSSCGRLKKAQGTNSPYPGHYPPQHXSTWNIEVENNOHVKVRFKFF 459
 QY 386 YLLEFGVPAGTCKDYVEINGEKYCGERSQFVVTSSNKNITVRFHSDQSYTDTGFLAEYL 445
 Db :|||||
 460 YLLEFGVPAGTCKDYVEINGEKYCGERSQFVVTSSNKNITVRFHSDQSYTDTGFLAEYL 519
 QY 446 SYDSSDPCPGQFTCTGR-CIRKELRCQGWADCTDHSDELNCSCDAGHQFTCKNFKCKPL 504
 Db :|||||
 520 SYDSSDPCPGQFTCTPHGAVIRKELRCQGWADCTDHSDELNCSCDAGHQFTCKNFKCKPL 579
 QY 505 FWVCDVNDGDNDEQGCSPAQTFRCNGKCLSKSQCNKGKDCGKDCGSDSEASCPCVNV 564
 Db :|||||
 580 FWVCDVNEGDNDEQGCICPAQTFRCNGKCLSKSQCNKGKDCGKDCGSDSEASCPCVNV 639
 QY 565 VTCTKHTYRCNLGLCLSKGNPECDGKDEKDCDGLRSFTRQARVVGTDADG 624
 Db :|||||
 640 VTCTKHTYRCNLGLCLSKGNPECDGKDEKDCDGLRSFTRQARVVGTDADG 699
 QY 625 EWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRFRYSFDPQWT-AFLGLHDQSO 683
 Db :|||||
 700 EWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRFRYSFDPQWT-AFLGLHDQSO 759
 QY 684 RSAP-GVOERRLKRISHPFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAG 742
 Db :|||||
 760 RPPWGCRRRLKRISHPFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAG 819
 QY 743 KAIWVTGHTQYGGTGALILQKGEIRVINQTTNENLLPQOITPRMCMVGLSGGVDSQC 802
 Db :|||||
 820 KAIWVTGHTQYGGTGALILQKGEIRVINQTTNENLLPQOITPRMCMVGLSGGVDSQC 879
 QY 803 GDSGGPLSSVEADGRIFQAGVWSWGDGCAQRNKPQVYTRLPFLFRDWIKENTGV 855
 Db :|||||
 880 GDSGGPLSSVEADGRIFQAGVWSWAGCVQRNKPQVYTRLPFLFRDWIKENTGV 932

Search completed: September 23, 2005, 13:09:19
 Job time : 92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:50:11 ; Search time 26 Seconds
(without alignments)
2454.806 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSDBARKGGGPKDFGAGL.....PGVYTRFLPFRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4681	100.0	855	US-09-027-337-2	Sequence 2, Appli
2	4681	100.0	855	US-09-644-600-2	Sequence 2, Appli
3	4681	100.0	855	US-09-654-600A-2	Sequence 2, Appli
4	3810	81.4	902	US-09-644-600-10	Sequence 10, Appl
5	3810	81.4	902	US-09-654-600A-10	Sequence 10, Appl
6	1319	28.2	241	US-09-657-986B-2	Sequence 2, Appli
7	703.5	15.0	798	US-08-200-900A-2	Sequence 2, Appli
8	703.5	15.0	798	US-08-794-042-2	Sequence 2, Appli
9	703.5	15.0	798	PCT-US94-00616-2	Sequence 2, Appli
10	692	14.8	1042	US-09-959-392-2	Sequence 2, Appli
11	664.5	14.2	1113	US-09-959-392-4	Sequence 4, Appli
12	588	12.6	407	US-09-734-675-4	Sequence 4, Appli
13	577	12.3	717	US-09-949-016-11182	Sequence 11182, A
14	574	12.0	699	US-09-949-016-6138	Sequence 6138, Ap
15	560.5	12.0	492	US-09-685-166A-895	Sequence 895, App
16	560.5	12.0	492	US-09-879-792-14	Sequence 14, Appl
17	560.5	12.0	492	US-09-679-426-895	Sequence 895, App
18	560.5	12.0	492	US-09-759-143-895	Sequence 895, App
19	558.5	11.9	492	US-09-759-143-934	Sequence 934, App
20	558.5	11.9	492	US-09-342-749-2	Sequence 2, Appli
21	558.5	11.9	492	US-09-691-840-2	Sequence 2, Appli
22	558.5	11.9	492	US-09-759-143-932	Sequence 932, App
23	558.5	11.9	510	US-09-949-016-11074	Sequence 11074, A
24	549.5	11.7	232	US-09-959-392-32	Sequence 32, Appl
25	547.5	11.7	235	US-08-944-483-65	Sequence 65, Appl
26	544.5	11.6	235	US-08-807-151-3	Sequence 3, Appli
27	544.5	11.6	235	US-09-478-957-3	Sequence 3, Appli

28	543.5	11.6	454	3	US-09-518-046-2	Sequence 2, Appli
29	533	11.4	521	4	US-09-949-016-11081	Sequence 11081, A
30	533	11.4	521	4	US-09-949-016-11082	Sequence 11082, A
31	533	11.4	521	4	US-09-949-016-11083	Sequence 11083, A
32	527	11.3	446	4	US-10-177-661-4	Sequence 4, Appli
33	521.5	11.1	477	4	US-10-177-661-6	Sequence 6, Appli
34	518.5	11.1	477	4	US-10-177-661-2	Sequence 2, Appli
35	518.5	11.1	562	4	US-09-879-792-12	Sequence 12, Appl
36	514.5	11.0	638	2	US-08-681-151-3	Sequence 3, Appli
37	512.5	10.9	283	3	US-08-807-151-1	Sequence 1, Appli
38	512.5	10.9	283	3	US-08-478-957-1	Sequence 1, Appli
39	511.5	10.9	418	1	US-08-508-448C-25	Sequence 25, Appl
40	511.5	10.9	418	4	US-09-370-838-82	Sequence 82, Appl
41	511.5	10.9	418	4	US-09-370-838-83	Sequence 83, Appl
42	511.5	10.9	418	4	US-09-854-133-82	Sequence 82, Appl
43	511.5	10.9	418	4	US-09-854-133-83	Sequence 83, Appl
44	510.5	10.9	418	4	US-09-370-838-62	Sequence 62, Appl
45	510.5	10.9	418	4	US-09-854-133-62	Sequence 62, Appl

ALIGNMENTS

RESULT 1
US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; OTHER INFORMATION: 23 to 2589 of Sequence 1
; Patent No. 5972616
US-09-027-337-2

Query Match	100.0%;	Score 4681;	DB 2;	Length 855;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 855;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGSDBARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKYKGGPGRWVLA	60	
Db	1	MGSDBARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKYKGGPGRWVLA	60	
Qy	61	VLIGLLVLIGIGFLVHQLQYRDVQKVFNGYMRITNFENFVDAYENSNSTFEVSLASKV	120	
Db	61	VLIGLLVLIGIGFLVHQLQYRDVQKVFNGYMRITNFENFVDAYENSNSTFEVSLASKV	120	
Qy	121	KDALKLYSGVPFLGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVEAEERVMABERVVM	180	
Db	121	KDALKLYSGVPFLGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVEAEERVMABERVVM	180	
Qy	181	LPFRARSLKSFVYTSVVAFPDTSKTQVORTODNCSFGLHARGVELMRTTFCGDPDPSYP	240	
Db	181	LPFRARSLKSFVYTSVVAFPDTSKTQVORTODNCSFGLHARGVELMRTTFCGDPDPSYP	240	
Qy	241	HARQWALRGDADSVLSLTFERSFDLASCDSRGSDLVTYNTLSPMEPHALVQLCGTTPPS	300	
Db	241	HARQWALRGDADSVLSLTFERSFDLASCDSRGSDLVTYNTLSPMEPHALVQLCGTTPPS	300	
Qy	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCCGRURKKAQGTNSPYYPGHY	360	
Db	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCCGRURKKAQGTNSPYYPGHY	360	

QY 361 PPNIDCTWNIIEVPNNQHVSKVFFYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 DB 361 PPNIDCTWNIIEVPNNQHVSKVFFYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRFHSDQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 DB 421 NSNKITVRFHSDQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNDEQSCPAQTFRCNSGKCLSK 540
 DB 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNDEQSCPAQTFRCNSGKCLSK 540
 QY 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 DB 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 QY 601 DDCGLRSFTRQARVVGTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 DB 601 DDCGLRSFTRQARVVGTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 DB 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 QY 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTGALILQKGEIRVINQTTCE 780
 DB 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTGALILQKGEIRVINQTTCE 780
 QY 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 DB 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855

RESULT 2
 US-09-644-600-2
 ; Sequence 2, Application US/09644600
 ; Patent No. 6451500
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hiroto
 ; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease
 ; FILE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/644,600
 ; CURRENT FILING DATE: 2000-08-23
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; PRIOR FILING DATE: 1999-10-20
 ; PRIOR APPLICATION NUMBER: 09/027,337
 ; PRIOR FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 2
 ; LENGTH: 855
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: TAGD-15
 US-09-644-600-2

Query Match 100.0%; Score 4681; DB 4; Length 855;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGSDDRKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVEKHGGRVVLAA 60
 DB 1 MGSDDRKGGGPKDFGAGLKYNSRHEKVNGLGEEVFLPVNNVKKVEKHGGRVVLAA 60
 QY 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

DB 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 QY 121 KDALKLLYSGVPFFLGPYHKSATVAFSEGSVIAYYWSFSIPQHLVEBAERVMAERVVM 180
 DB 121 KDALKLLYSGVPFFLGPYHKSATVAFSEGSVIAYYWSFSIPQHLVEBAERVMAERVVM 180
 QY 181 LPPRARSLSKSPVVTSSVAFPTDSKITVQRTQDSCSFGHLHARGVELMRFTTGGFPDPS 240
 DB 181 LPPRARSLSKSPVVTSSVAFPTDSKITVQRTQDSCSFGHLHARGVELMRFTTGGFPDPS 240
 QY 241 HARQWALRGADSVLSITFRSPDLASCDERGSDLVTYNTLSMPHEHALVOLCCTYPPS 300
 DB 241 HARQWALRGADSVLSITFRSPDLASCDERGSDLVTYNTLSMPHEHALVOLCCTYPPS 300
 QY 301 YNLTFHSSQNVLLTLITNTERRHPPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
 DB 301 YNLTFHSSQNVLLTLITNTERRHPPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
 QY 361 PPNIDCTWNIIEVPNNQHVSKVFFYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 DB 361 PPNIDCTWNIIEVPNNQHVSKVFFYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRFHSDQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 DB 421 NSNKITVRFHSDQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNDEQSCPAQTFRCNSGKCLSK 540
 DB 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNDEQSCPAQTFRCNSGKCLSK 540
 QY 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 DB 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 QY 601 DDCGLRSFTRQARVVGTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 DB 601 DDCGLRSFTRQARVVGTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 DB 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 QY 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTGALILQKGEIRVINQTTCE 780
 DB 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTGALILQKGEIRVINQTTCE 780
 QY 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 DB 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855

RESULT 3
 US-09-654-600A-2
 ; Sequence 2, Application US/09654600A
 ; Patent No. 6649741
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hiroto
 ; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease
 ; FILE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/654,600A
 ; CURRENT FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; PRIOR FILING DATE: 1999-10-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 2

; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2

Query Match 100.0%; Score 4691; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDAGAGLYNRSHEKNGLEEGVEFLPVNNKKVEKHGGRWVLLA 60
DB 1 MGS DRARKGGGPKDAGAGLYNRSHEKNGLEEGVEFLPVNNKKVEKHGGRWVLLA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
DB 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120

QY 121 KDALKLLYGVPPFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
DB 121 KDALKLLYGVPPFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180

QY 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTGPFPSPYPA 240
DB 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTGPFPSPYPA 240

QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
DB 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSGCCGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSGCCGLRKAQGTFFNSPYYPGHY 360

QY 361 PNIDCTWNIENVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
DB 361 PNIDCTWNIENVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
DB 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480

QY 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540
DB 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SOQCNGKDDCGDSDGDEASCPCNVVTCYKHYRCLNGLCLSKGNPECDGKDCSDGDEK 600
DB 541 SOQCNGKDDCGDSDGDEASCPCNVVTCYKHYRCLNGLCLSKGNPECDGKDCSDGDEK 600

QY 601 DCDCLGSRSTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DCDCLGSRSTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNFDTFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNFDTFDYDIALLELEKP 720

QY 721 ABYSNMVRPCLPDAASHVPPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780
DB 721 ABYSNMVRPCLPDAASHVPPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780

QY 781 POQITPRMVCVFLSGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
DB 781 POQITPRMVCVFLSGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840

QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-644-600-10

Query Match 81.4%; Score 3810; DB 4; Length 902;
Best Local Similarity 81.4%; Pred. No. 1.1e-270;
Matches 686; Conservative 71; Mismatches 86; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDAGAGLYNRSHEKNGLEEGVEFLPVNNKKVEKHGGRWVLLA 60
DB 1 MGS DRARKGGGPKDAGAGLYNRSHEKNGLEEGVEFLPVNNKKVEKHGGRWVLLA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
DB 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120

QY 121 KDALKLLYGVPPFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
DB 121 KDALKLLYGVPPFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180

QY 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTGPFPSPYPA 240
DB 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTGPFPSPYPA 240

QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
DB 241 HARCOWALRGDADSVLSLTFRSFDVASPCDEHSDLVTVYNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSGCCGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPMSGCCGLRKAQGTFFNSPYYPGHY 360

QY 361 PNIDCTWNIENVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
DB 361 PNIDCTWNIENVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
DB 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480

QY 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540
DB 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SOQCNGKDDCGDSDGDEASCPCNVVTCYKHYRCLNGLCLSKGNPECDGKDCSDGDEK 600
DB 541 SOQCNGKDDCGDSDGDEASCPCNVVTCYKHYRCLNGLCLSKGNPECDGKDCSDGDEK 600

QY 601 DCDCLGSRSTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DCDCLGSRSTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNFDTFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNFDTFDYDIALLELEKP 720

Db 661 DKNFKYSDYTWMTAFGLLDQSRSSAGVQELKRIITHPSFNDFDYDIALLELEKS 720
QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGLALILQKGEIRVINQTCENLL 780
Db 721 VEYSTVVRPCLPDATHVFPAGKAIWVTGHTKEGGTGLALILQKGEIRVINQTCEDLM 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSVSEADGRIFQAGVSVSGDGCAGRNKPGVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSAEGKGRMFQAGVSVSGEGCAQRNKGVT 840
QY 841 RLP 843
Db 841 RLP 843

RESULT 5

US-09-654-600A-10
; Sequence 10, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: Overexpressed in Carcinomas
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-654-600A-10

Query Match 81.4%; Score 3810; DB 4; Length 902;
Best Local Similarity 81.4%; Pred. No. 1.1e-270;
Matches 686; Conservative 71; Mismatches 86; Indels 0; Gaps 0;
QY 1 MGSDRKGGGGKDFGAGLKYNRHEKVNLEEGVEFLPVANNVKKVKGPRVWLAA 60
Db 1 MGSNRKKGAGGSDDFGAGLKYSRLNNGVEEGVEFLPVANNVKKVKGPRVWLVA 60
QY 61 VLIGLLVLLGIFLYVHLQYRDVRVQKVFNGYMRITNENFVDAYENSSTFEVSLASKV 120
Db 61 VLFSFLLLSLMAELLVWHFHYRNVQKVFNGHLRITNEIFLDAYENSTSTFEVSLAQV 120
QY 121 KQALKLLYSGVPLPYHESAVTAPSEGSVAYYWSSEFISQHLVEAEERVAERVVM 180
Db 121 KEALKLLYNEVPVLPYHKKSAVTAPSEGSVAYYWSSEFISQHLVEAEERVAERVVT 180
QY 181 LPPRARSLSFVYTSVVAFPTDSKTQRTQDSCSFLHARGVELMRFTTPGFPDPSYPA 240
Db 181 LPPRARKLSFVYTSVVAFPIDRMLQRTQDSCSFLAHAGAAVTRFTTPGFPNSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHVALVOLCTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFRSFDVAPCDEHGSDDLVTYVDSLSPHVAVRLCGTFFSPS 300
QY 301 YNLTFHSSQNVLLITLITTRRHQGFATFPQLPRMSSCGRLRKAQCTFNSPYPYGHY 360
Db 301 YNLTFHSSQNVLLITLITNTRHGLGFATFPQLPRMSSCGVGLSDTQCTFNSPYPYGHY 360
QY 361 PNINDCTWNIENPVNNOHVKSFKFFVLLPBPVAGTCPKDYVEINKEKYCGRSQFVYTS 420
Db 361 PNINCTWNIKVPNNRNVKRFKLFYLDVDFNVPGSCTKDYVEINKEKSGRSQFVYSS 420

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSDSPCPGQQTCTRGRCIRKELRCDGWADCTDH 480
Db 421 NSSKITVRFHSDHSYTDGFLAEYLSYSDSNPCPCGMFMCKTGRCIRKELRCDGWADCPDY 480
QY 481 SDELNCSADAGHOFCKKPKFKPLFWVCDSDVNDGDSDEQCCSCPAOTFRCSNGKCLSK 540
Db 481 SDERYCRNATHQFTCKNQFCFKPLFWVCDSDVNDGDSDEGSCSPAGSFKCSNGKCLPQ 540
QY 541 SOQCNKGDDCGDSDSEASCPCVNVVTCFKHYRCLNGLCLSKNGNPECDGKEDCSGSDSK 600
Db 541 SQKCNKNDKCGDSDSEASCDSNVVSTKYRRCQNGLCCLSKNGNPECDGKEDCSGSDSK 600
QY 601 DCDCLRSFTRQARVVGCTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 NCDGLRSFTRQARVVGCTNADEGEWPMQVSLHALGQGHICGASLISPDWLVSAAHCFOD 660
QY 661 DRGFRYSDPTOWTAFGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKP 720
Db 661 DKNFKYSDYTWMTAFGLLDQSRSSAGVQELKRIITHPSFNDFDYDIALLELEKS 720
QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGLALILQKGEIRVINQTCENLL 780
Db 721 VEYSTVVRPCLPDATHVFPAGKAIWVTGHTKEGGTGLALILQKGEIRVINQTCEDLM 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSVSEADGRIFQAGVSVSGDGCAGRNKPGVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSAEGKGRMFQAGVSVSGEGCAQRNKGVT 840
QY 841 RLP 843
Db 841 RLP 843

RESULT 6

US-09-657-986B-2
; Sequence 2, Application US/09657986B
; Patent No. 6797504
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Araldi, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matriptase or
; TITLE OF INVENTION: MTSP1
; FILE REFERENCE: Corvas 255/049
; CURRENT APPLICATION NUMBER: US/09/657,986B
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-657-986B-2

Query Match 28.2%; Score 1319; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 6.7e-89;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 615 VVGCTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTOWTA 674
Db 1 VVGCTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTOWTA 60
QY 675 FLGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCLCPD 734
Db 61 FLGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKPAEYSSMVRPCLCPD 120
QY 735 ASHVFPAGKAIWVTGHTQYGGTGLALILQKGEIRVINQTCENLLPOQITPRMCMVGL 794
Db 121 ASHVFPAGKAIWVTGHTQYGGTGLALILQKGEIRVINQTCENLLPOQITPRMCMVGL 180

QY 795 SGGVDSQCGSGPLSSVEADGRIFQAGVVSMDGCAQRNKGQVYTRLPFRDWMKENTG 854
Db 181 SGGVDSQCGSGPLSSVEADGRIFQAGVVSMDGCAQRNKGQVYTRLPFRDWMKENTG 240
QY 855 V 855
Db 241 V 241

RESULT 7

US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinhert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 15.0%; Score 703.5; DB 1; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPRMSSCGG--RLRKAQGTFSPPYGHYPNIDCTNIEVPNNQHVKSFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLMEPTTFTSINFPNSYPNQAFCIWNLNAQKGNQLHFEFDLENI 355
QY 391 GVPAGTCPKDYVEI--NGEKYCGERSQFV-----VTSNSNKITVRFHSDQSYTDT 438
Db 356 A-----DVEIRDEG---GDSLFLAVYTGPGFVNDVFTNRMVTLFTIDNMLAQ 404
QY 439 GLAEVLS---YDSDPC--PGQFTCTGTCIRKELRCGWDCTDHSDELNC-----SC 488
Db 405 GFKANFTTGYGLGIPCEKEDNFQCKDGCIPLVNLCDFGPHCKDGSDEAHCVRFLNFGTT 464
QY 489 DAGH--QFTCKNFKCKPLFWVCDSDNDSCDNGSCPAOTFRCSNGKCLKSKSQOQNG 546
Db 465 DSSGLVQFRIQS-----LWHV-----ACAE-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGDGSDEAC-----PKVNVVTCRTHYRCNLGCLSKGNPECDGKEDCSGSDK 600
Db 500 ---LGTGNSVFPFGTGGGPGVYVNLNAP-----NGSLILTPSQQC-----LEDSLILL 544

QY 601 DCD-----CGLRSFTROA--RVVGGTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNVKSCKGLVTVQSVSPKIVGGSDSREGAWPMVVALYFDDQ--QVCGASLVSRLMLVSA 603
QY 656 HCYIDDRGRFYSDPTQWTAFLGLHDQSORSAPGVQERRLKRILSHPPFFNDFTFDYDIAL 715
Db 604 HCVYG-----RNMEPSKWKAVLGLHNASNLTSQIETRLIDQIVINPHYNKRKNNDIAM 659
QY 716 ELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGHTQYGGTGCTGALILQKGRIRVINQTT 775
Db 660 HLEMKVNTDYIOPICLPEENQVFPFGRICSIAGHGLIYQGSTADVLQEAADVPLLSNEK 719
QY 776 CENLLPQ--QITPRMCMVGLSGVDSCQDGGPLSSVEADGRIFQAGVVSMDGCAQRN 834
Db 720 CQOQMPYNTNITENMVCAGYEAGGVDSQDGGPLMCQE--NNRWLLAGVTSFGYOCALPN 778
QY 835 KPGVYTRLPFRDWMK 850
Db 779 RPGVYARVPRTWIIQ 794

RESULT 8

US-08-794-042-2
; Sequence 2, Application US/08794042
; Patent No. 6746859
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/794,042
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/200,900
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinhert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-794-042-2

Query Match 15.0%; Score 703.5; DB 4; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPRMSSCGG--RLRKAQGTFSPPYGHYPNIDCTNIEVPNNQHVKSFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLMEPTTFTSINFPNSYPNQAFCIWNLNAQKGNQLHFEFDLENI 355
QY 391 GVPAGTCPKDYVEI--NGEKYCGERSQFV-----VTSNSNKITVRFHSDQSYTDT 438

Db 356 A-----DVAIRIDGE---GDSLFLAVYTGPGVNDVFTTNRMTVLFITDNMLAQ 404
QY 439 GFLAEYLS---YDSSDPC-PGQFCTGRCIRKELRCDCGWADCTDHSDELNC-----SC 488
Db 405 GFKNFTTGYGLGIPBCKEDNFQCKGECIPLVNLCDGPFHCKGSDAHCVRLEPFGTT 464
QY 489 DAGH--OFTCKNFKCKPLFWVCDSDVNDGDSDEQSCCPAQOTFRCSNGKCLSKSQCCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEN-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGSDGEASC-----PKNVVTCYKHTYCLNGLCLSKGNPECDKEDCSGSDK 600
Db 500 ---LGTGNSVPTFTGGGPPVNLNTAP-----NGSLILTPSOQC-----LEDLILL 544
QY 601 DCD---CGLRSFTROA--RVVGGTDADGEGWPMQVSLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNYKSCGKLVQEVSPKIVGGSDREGAPWVVALYFDDQ-QVCGASLVRDMLVSA 603
QY 656 HCVIDRGRFYSPTQWTAFLGLHDOSORSAPGQVBRRLKRIISHPPFNDFTFDYDIAL 715
Db 604 HCVYG---RNMFPKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRKNNDIAMM 659
QY 716 ELEKPAEYSSMRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTT 775
Db 660 HLEKMYNTDYIQCIPLEENQVFPFGRICSIAGWGALYIYQGSTADVLQADVPILSNEK 719
QY 776 CENLLPQ-QITPRMCMVGLSGVDSQCGSDGGPLSSVEADGRIFOAGVWSMGDCAQRN 834
Db 720 CQOQMEYNITENMVCAGYEAGGVDSQCGSDGGPLMCQE-NNRWLLAGVTSFGYQCALPN 778
QY 835 KPGVYTRLPFRDWIK 850
Db 779 RGVYARVPRFTWIIQ 794

RESULT 9

PCT-US94-00616-2

; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: PCT/US94/00616
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

PCT-US94-00616-2

Query Match 15.0%; Score 703.5; DB 5; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPMSSCGG--RLRKAQTFNSPYPGHYPPNIDCTNIEVPNNQHVKVSFKPYLLEP 390
Db 298 ELP--TDCGGPDLMEPNTFTTSINFPNSYPNQAFCIMNLNAQKQNLQHFQEPDLENI 355
QY 391 GVPAGTCPKDYVEI--NGEKYCCERSQFV-----VTSNSKNTVTRFHSQSYTDT 438
Db 356 A-----DVAIRIDGE---GDSLFLAVYTGPGVNDVFTTNRMTVLFITDNMLAQ 404

QY 439 GFLAEYLS---YDSSDPC-PGQFCTGRCIRKELRCDCGWADCTDHSDELNC-----SC 488
Db 405 GFKNFTTGYGLGIPBCKEDNFQCKGECIPLVNLCDGPFHCKGSDAHCVRLEPFGTT 464
QY 489 DAGH--OFTCKNFKCKPLFWVCDSDVNDGDSDEQSCCPAQOTFRCSNGKCLSKSQCCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEN-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGSDGEASC-----PKNVVTCYKHTYCLNGLCLSKGNPECDKEDCSGSDK 600
Db 500 ---LGTGNSVPTFTGGGPPVNLNTAP-----NGSLILTPSOQC-----LEDLILL 544
QY 601 DCD---CGLRSFTROA--RVVGGTDADGEGWPMQVSLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNYKSCGKLVQEVSPKIVGGSDREGAPWVVALYFDDQ-QVCGASLVRDMLVSA 603
QY 656 HCVIDRGRFYSPTQWTAFLGLHDOSORSAPGQVBRRLKRIISHPPFNDFTFDYDIAL 715
Db 604 HCVYG---RNMFPKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRKNNDIAMM 659
QY 716 ELEKPAEYSSMRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTT 775
Db 660 HLEKMYNTDYIQCIPLEENQVFPFGRICSIAGWGALYIYQGSTADVLQADVPILSNEK 719
QY 776 CENLLPQ-QITPRMCMVGLSGVDSQCGSDGGPLSSVEADGRIFOAGVWSMGDCAQRN 834
Db 720 CQOQMEYNITENMVCAGYEAGGVDSQCGSDGGPLMCQE-NNRWLLAGVTSFGYQCALPN 778
QY 835 KPGVYTRLPFRDWIK 850
Db 779 RGVYARVPRFTWIIQ 794

RESULT 10

US-09-959-392-2

; Sequence 2, Application US/09959392
; Patent No. 6806075
; GENERAL INFORMATION:
; APPLICANT: MORSE, MICHAEL JOHN
; APPLICANT: WU, QINGYU
; APPLICANT: YAN, WEI
; TITLE OF INVENTION: CORIN, A SERINE PROTEASE
; FILE REFERENCE: BERLX 74A
; CURRENT APPLICATION NUMBER: US/09/959,392
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: PCT/EP99/03895
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 09/092,029
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 09/314,967
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1042
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-959-392-2

Query Match 14.8%; Score 692; DB 4; Length 1042;
Best Local Similarity 27.8%; Pred. No. 3.8e-42;
Matches 210; Conservative 98; Mismatches 252; Indels 196; Gaps 35;
QY 202 DSKTVQRTDNCSGFLHARGVELMRTTPGFPDPSYPNAHRCQWALRGDAD-----253
Db 365 DHDCVDSKDEVNCS--CHSQGLVECR-----NQCCIPSTFQCD-----GDECDKGSDEE 412
QY 254 --SVLSLTFRSFD-----LASCDERGSDLVTVYNTL-----SPMEPHAVVOLCGTYPPS 300
Db 413 NCSVIQTSCEGQDQRCLYNPCLDSCG--GSSLCDPNNSLNKSCQCEPITL--ELCWNLP-- 467
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFETFFQLPRSSCCGRLRK-AQGTFFNFPYPGH 359

Db 468 YNST-----SYPNY-----FGHRTQKEASISWESSLPAL 497
QY 360 YPNIDICTWNIEVNNQHVKSFKFYLLPEPGVAGTCKPKDYVEIN-GEK-----Y 409
Db 498 VQTN-----CYKLMFFSCTILVPKC-----DVTGERIPPCRALCEH 535
QY 410 CGRSQFVV-----TSNSNKITVRPHSDQSYDTGFLA-EYLSYDSSDPC-PGQF 457
Db 536 SKERCESVLGIVGLWPEDTDCS-----QPPEENSQNTCLMPDEVV-----ECCSPSHF 585
QY 458 TCHTGCIRKELCDGWADCTDHSDELNCSDAGHQTFC-KNFKCPLFWVCDSDVNDGSD 516
Db 586 KCSGQCVLASRCDDQADDDSDDEENGCKERDLWECPNKCQCLKHTVICDGFPPDCPD 645
QY 517 NSDEQCS-CPAQTFRCNGKCLSKSQOQNGKDDCGDGDSEASCPKVV-----564
Db 646 YNDEKNCSPQDDELECANHACVSRDLWCDGEADCSDSDEWDCVTLINVNSSSFLMVH 705
QY 565 -----VTC-----TKHT-YRLNGLCL 580
Db 706 RAATEHHVCADGHOEILSQLACKOMGLGEPSTVKLIOEQEKPRWLTLSHNWESLNGTTL 765
QY 581 SK---GNPCDGEDSDSDXDCDGLRSFTR-QARVVGGTDADEGEWPMQVSLHALG 636
Db 766 HELLVNGQSCESKSLILCTKO--DCGRPAARMNKRILGRTSRPGRWPMQCSLQSEP 823
QY 637 QGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQSORAPGVOERLKR 696
Db 824 SGHICCVLIACKWLTVAHCF-EGR-----ENNAWKVVLGINLNDHPSV-FMOTRFVKT 877
QY 697 IISHPPFNDFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYG 756
Db 878 IILHPRYSRAVVDYDISIVELSDSETGVPRVCLPNPEQWLEPDTYCVITGWGH--WG 935
QY 757 GTGALLIQGEIRVINTTCENLLPQO-ITPRMVCVGLSGVDSQCGSDGGLSSVEAD 815
Db 936 NKMPFKLQGEVRILSLHCQSYFDMKTTITRMICAGYESGTVDSCMGSDGGLVCEKPG 995
QY 816 GRIFOAGVSWGDGCAOR-NKPGVYTRLPFRDWIK 850
Db 996 GRWTLFGLTWSGSCVFSKVLGPGVYSNVSYFVWEIK 1031

RESULT 11

US-09-959-392-4

; Sequence 4, Application US/09959392

; Patent No. 6806075

; GENERAL INFORMATION:

; APPLICANT: MORSE, MICHAEL JOHN

; APPLICANT: YAN, WEI

; APPLICANT: YAN, WEI

; TITLE OF INVENTION: CORIN, A SERINE PROTEASE

; FILE REFERENCE: BERLX 74A

; CURRENT APPLICATION NUMBER: US/09/959,392

; CURRENT FILING DATE: 2001-10-25

; PRIOR APPLICATION NUMBER: PCT/EP99/03895

; PRIOR FILING DATE: 1999-06-04

; PRIOR APPLICATION NUMBER: 09/092,029

; PRIOR FILING DATE: 1998-06-05

; PRIOR APPLICATION NUMBER: 09/314,967

; PRIOR FILING DATE: 1999-05-20

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 1113

; TYPE: PRT

; ORGANISM: Mus sp.

US-09-959-392-4

Query Match

Best Local Similarity 14.2%; Score 664.5; DB 4; Length 1113;

Matches 176; Conservative 78; Mismatches 221; Indels 115; Gaps 22;

QY 352 NSPYTPGHYP-----PNIDCTWNIE-----VNNQHVKSFKFYLLPEPGVAGTCTP 398
Db 533 NLPYNHTYPNYLGHRTQKEASISWESSLPALVQNTCYKLMFFACTILVPKCDVNT-- 590
QY 399 KDYVINGEK-----YCGERSQFVV-----TSNSNKITVRPHSDQSYDTGTF 440
Db 591 -----GQRIPPCRLLCESKERCESVLGIVGLWPEDTDCN-----QPPESSDQNTCL 639
QY 441 LAEYLSYDSSDPCQOFTCTGRCIRKELCDGWADCTDHSDELNCSDAGHQTCTCK-NK 499
Db 640 LP---NEDEVESPSHFKCRSGRCVILGSRCDGQADDDSDDEENGCKERDLWECPFNK 696
QY 500 FCKPLFWVCDSDVNDGSDDEQCS-CPAQTFRCNGKCLSKSQOQNGKDDCGDGDSEAS 558
Db 697 QCLKHTLICDGFPPDCPDSDMDEKNCSCQDNELECANHECVPRDLWCDGWVDCSDSDWEG 756
QY 559 CPKVN-----VVTCTKTYRCLNG-----LC-----LSKGNPECDGKE 591
Db 757 CVTLKNGNSGSLTLVHKSACEHHVCADGWRETLSQLACKOMGLGEPSTVKLIPIGQEQO 816
QY 592 -----DCSDGSDKED-----C---DCGLRSFTR-QARVVGGTDDAD 622
Db 817 WURLYPNNENLNGSTLQELLVYRHSCPSRSEISLCSKQDCGRRPAARMNKRILGRTSR 876
QY 623 EGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQ 682
Db 877 PGRWPMQCSLQSEPESHICGCVLIACKWLTVAHCF-EGR-----EDADVWKVVLGINLND 931
QY 683 QRSAPGVOERLKRILSHPPFNDFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 742
Db 932 HPSG-FMOTRFVKTILLHPRYSRAVVDYDISIVELSDINETSYYRVPVCLPSPPEYLEPD 990
QY 743 KAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQO-ITPRMVCVGLSGVDS 801
Db 991 TYCITITGHH--MGNMPEFKLQGEVRILPLEQCSYFDMKTTITRMICAGYESGTVDSC 1048
QY 802 QGDSGGLSSVEADGRIFQAGVSWGDGCAOR-NKPGVYTRLPFRDWIK 850
Db 1049 MGDSGGLVCPERPQGWTLFGLTWSGSCVFSKVLGPGVYSNVSYFVGMIE 1098

RESULT 12

US-09-734-675-4

; Sequence 4, Application US/09734675

; Patent No. 6365391

; GENERAL INFORMATION:

; APPLICANT: WEBSTER, Marion et al

; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL000862

; CURRENT APPLICATION NUMBER: US/09/734,675

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 407

; TYPE: PRT

; ORGANISM: Human

US-09-734-675-4

Query Match 12.6%; Score 588; DB 3; Length 407;

Best Local Similarity 45.9%; Pred. No. 4.7e-35;

Matches 119; Conservative 38; Mismatches 82; Indels 20; Gaps 6;

QY 604 CGLR---SFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYTD 660

Db 162 CGTRRSKTLQSLRIVGGTEVEGEWPMQASLQWDG-SHRCGATLINATLWLSAAHCF- 219

QY 661 DRGFYSDPTQWTAFLGLHDQSORAPGVOERLKRILSHPPFNDFDYDIALLELEK 720

Db 220 ----TYKNPARWTASFGV-----TIKPSMKRGLRRIIVHEKYPKSHDYDISLAELSSP 270

```
QY 721 ABYSSMVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTGALILQKGEIRVINOTTCENLL 780
Db 271 VPTNAVHRVCLPDASYEQFGDVMFTGFGALKNDGYSQHLRQAQVTLIDATTGNE-- 328
QY 781 PQ----QITPRMVCVGLSGVDSGCGDGLSSVEADGRIIFQAGVWSGDCACARNKP 836
Db 329 PQAYNDAITPRMLCAGLEBKTACQDGGPLVSSDARDIWIYLAGIVSGDSCAKPNKP 388
QY 837 GVTYTRLPFLFRDMIKENTGV 855
Db 389 GVTYTRVALRDMITSKTI 407

RESULT 13
US-09-949-016-11182
; Sequence 11182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11182
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11182

Query Match 12.3%; Score 577; DB 4; Length 717;
Best Local Similarity 25.3%; Pred. No. 6.2e-34;
Matches 193; Conservative 110; Mismatches 269; Indels 192; Gaps 31;

QY 203 SKTVQRTQDNCSCFGLHARGVELMRFTTPGPPDSPYPAHARCOWALRGDADSVLSITFRS 262
Db 33 SKASHTVELNNMFG-----QIQSPGYPDS-YPSDSEVTWNITVPDGERIKLYFMH 82
QY 263 FDLASCDERGSDLVTVNTLSPMEPHALVQLCG-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 137
Db 83 FNLESSYLCEYDYVKV-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 137
QY 307 SSQNVLLITLITNTERRHGFEATFFQLP-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 335
Db 138 SD-----FSENERFTGFDHYMAVDVDECKERDEBELSCDHCHYCHNYIGGYCSCRF 188
QY 336 -----RMSCCGRLRKAQGTFSNYPGHPYPPNIDCTWNIENPVNNOHVKSFKF 384
Db 189 GYLHTDNRTCRVECDNLFTQRTGVITSPDPFPNPKSSECLYTIIELEGFWNLQPED 248
QY 385 FYLLE--PGVPAGTCKPKDYVEIN-GEK-----YCGERSQFVVTNSNKTIVRPHSDQSYTD 437
Db 249 IFDIEDHPEVP--CPDYIKIKVGPVKVLGPFPGCEKAPBIPITQSHSVLILFHSDNSGEN 305
QY 438 TGFLAELYSVDSDDPCP-----GQFTCRGTCIRKELRCDCGWADCTDHSDELNCSDA 490
Db 306 RGWRLSYRA--AGNECEPQLPPVHGKLEPSQAKYFPK-----DOVLVSCUT 349
QY 491 GHQF-----TCRKNKCKPLFW-----VCSVNDGNSDEQGCSCPAQ-----TFR 531
Db 350 GYKVLKDNVEMDTFQIECLDKGTWSNKIPTCKIVD-----CRAPGELEHGLITFS 399
QY 532 CSNGKCLSSQCGKXGDCGSDGDEASCPCNVVVTCTKTYRC-----LNGICLSKGNP 585
Db 400 TRNNLTYYKSEI---KYSQOE-----PYKMLNNTGIYTCSAQGVMMNKV-LGRSLP 448
```

```
QY 586 ECDGKEDCSDGSDKDCDCCGLRSFTRO--ARVVGTTDADEGEWPMQVSLHALCOGHICGA 643
Db 449 TCLPV-----CGLPKFSKLMARIFNGRPAQKGTTPMTAMLSHLNGQPFQCG 495
QY 644 SLISPNWLVSAAHCY-----IDDRFRYSD-----PTQWTAFLGLHDOSQSAPOGVQERRLK 695
Db 496 SLUGSSWIVTAACHLHQSLDPDPTLRSDLLSPSDFKIILGKH-WRLASDENEQHLGVK 554
QY 696 RIISHPFNDFDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGNGHTQY 755
Db 555 HTTLHPQYDPDNTFENDVALVELLSPVLNAFWNPICLPEGPQ--QEGAMVIVSGW-KQF 611
QY 756 GGTGALILQKGEIRVINOTTCENL---LPQQITPRMVCVGLSGVDSGCGDGLSSV 812
Db 612 LQRPPELTMEIPIVDHSTCQKAVAPLKCKVTRDMICAGEKEGKDCACAGDSGGGPMVTL 671
QY 813 EAD-GRIFQAGVWSGDCGCAQBNKPGVYTRLRPLFRDWIKENTGV 855
Db 672 NRERGQWLVGTVSGDDCGKRDYGVSYIHHNKDWIQRVTGV 715

RESULT 14
US-09-949-016-6138
; Sequence 6138, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6138
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6138

Query Match 12.3%; Score 574; DB 4; Length 699;
Best Local Similarity 25.3%; Pred. No. 1e-33;
Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;

QY 203 SKTVQRTQDNCSCFGLHARGVELMRFTTPGPPDSPYPAHARCOWALRGDADSVLSITFRS 262
Db 15 SKASHTVELNNMFG-----QIQSPGYPDS-YPSDSEVTWNITVPDGERIKLYFMH 64
QY 263 FDLASCDERGSDLVTVNTLSPMEPHALVQLCG-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 306
Db 65 FNLESSYLCEYDYVKV-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 119
QY 307 SSQNVLLITLITNTERRHGFEATFFQLP-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGFSMISITPR 335
Db 120 SD-----FSENERFTGFDHYMAVDVDECKERDEBELSCDHCHYCHNYIGGYCSCRF 170
QY 336 -----RMSCCGRLRKAQGTFSNYPGHPYPPNIDCTWNIENPVNNOHVKSFKF 384
Db 171 GYLHTDNRTCRVECDNLFTQRTGVITSPDPFPNPKSSECLYTIIELEGFWNLQPED 230
QY 385 FYLLE--PGVPAGTCKPKDYVEIN-GEK-----YCGERSQFVVTNSNKTIVRPHSDQSYTD 437
Db 231 IFDIEDHPEVP--CPDYIKIKVGPVKVLGPFPGCEKAPBIPITQSHSVLILFHSDNSAEN 287
QY 438 TGFLAELYSVDSDDPCP-----GQFTCRGTCIRKELRCDCGWADCTDHSDELNCSDA 490
```

Db 288 RGNLSYRA--AGNECEPQPPVHGKIEPSQAKYFFK-----DQVLVSCDT 331
Qy 491 GHQF-----TCKNFKCKPLFW-----VCSVNDGDSNDEQGSCEPAQ-----TFR 531
Db 332 GYKVLKDNVEMDTFQIECLKDGTWSNKIPTCKIVD-----CRAPGELEHGLITPS 381
Qy 532 CSNGKCLSKSQOQNGKDDCGSDSDEASCPKVNVTCTKHYRC-----LNGLCCLSKGNP 585
Db 382 TRNNLTYYKSEI---KYSQOE-----PYKMLNNNTGIYTCSAQGVMMKV-LGRSLP 430
Qy 586 ECDGKEDCSDGSDKDCGLRSFTQ--ARVVGTTDADGEWPMQVSLHALQGGHICGA 643
Db 431 TCLPV-----CGLPKSRKLMARFNFRPAOKGTPMTAMLHSLNGOPFCGG 477
Qy 644 SLISPNLVSAAHY---ID--DRGFRYSQ--PTOWTAFGLGHQDSQRSAPGVERRLK 695
Db 478 SILGSSWIVTAACHLQSLDQDPTLRDSDLSPDFKILLKH-WRLSDENEQHLGVK 536
Qy 696 RIISHFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQY 755
Db 537 HTTLHPQYDNPFPENDVALVELLESPLNAFVMPICLPEGPQ--QEGAMVIVSGWG-KQF 593
Qy 756 GTGALILQGEIRVINQTTCNL---LPQOITPRMCMVGLSGVDSCQSDSGPLSSV 812
Db 594 LQRFPELTLEIPIIVDHTCQKAYAPLKKVTRDMICAGEKEGGKDACAGDSGGPMVTL 653
Qy 813 EAD-GRIFOAGVYVSWGDCAQRNKPQVYTRPLFRDRIKENTGV 855
Db 654 NREGQWYLVGVYSGDDCGKDKORYGVSYIHHNKDWIORVTGV 697

RESULT 15

US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Mark W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepier, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 12.0%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 6.2e-33;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

Qy 522 GCSCEPAQTFR--SNKCLSKSQOQNGKDDCGSDSDEASCPKV---NVV----- 565
Db 110 GSKCSNGIECDSSGTCINPSNWCDSVSHCPGGBDENRCVRLYGNFILIOMYSQKSWH 169

Qy 566 -----TCT-----KHYRCLNG-----LCLSKGNPECDGKEDCSD 595
Db 170 PVCQDDWNNYGRAACRDMGYKNNFYSSQGIYDDSGSTSPMKLNTSAGNVDIYKKLYHSD 229
Qy 596 GSDEK-----DCDCGLRSFTQARVVGTTDADGEWPMQVSLHALQGGHICGASLIS 647
Db 230 ACSSKAVVSLRCLACGVNLNS-SRQSRIVGGESALPGAWPMQVSLH-VQNVHVCGGSIIT 287
Qy 648 PNWLVSAAHCYIDDRGFYSPTOWTAFGLGHQDS-QRSAPGVQERRLKRILISHPPFNDP 706
Db 288 PEWIVTAACHVEKP-----LNNPWHMTAFAGILRQSFMYGAGYQ---VQKVISHPNYDSK 340
Qy 707 TFDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGCALILQKG 766
Db 341 TKNDIALMLKLOKPLTFNDLVKPVCLPNPMMLOPEQLCMISGWGATEBKGTSEVLNAA 400
Qy 767 EIRVINQTTCN--LLPQOITPRMCMVGLSGVDSCQSDSGPLSSVEADGRI--PQAGV 823
Db 401 KVLITETQRCNSRYVYDNLITPAMICAGFLQGNVDSQCGDSGGPL--VTSNNNIWWLIGD 458
Qy 824 VSWGDDCAQRNKPQVYTRPLFRDRI 849
Db 459 TSWGSGCAKAYRPGVYGNVMVFTDWI 484

Search completed: September 23, 2005, 12:59:40
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:56:11 ; Search time 174 Seconds
(without alignments)
1999.974 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSDRARKGGGPKDFGAGL.....PGVYTRFLPRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1826521 seqs, 407012169 residues

Total number of hits satisfying chosen parameters: 1826521

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:**
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4681	100.0	855	10	US-09-776-191-2
2	4681	100.0	855	14	US-10-099-700A-2
3	4681	100.0	855	14	US-10-190-030B-2
4	4681	100.0	855	14	US-10-302-840A-2
5	4681	100.0	855	14	US-10-267-219-2
6	4681	100.0	855	14	US-10-112-221A-2
7	4681	100.0	855	14	US-10-144-271-2
8	4681	100.0	855	15	US-10-147-211A-2
9	4681	100.0	855	15	US-10-156-214A-2
10	4681	100.0	855	15	US-10-072-012-352
11	4681	100.0	855	15	US-10-072-012-411

12	4681	100.0	855	15	US-10-072-012-418
13	4681	100.0	855	15	US-10-600-187-2
14	4681	100.0	855	17	US-10-612-466B-2
15	4676	99.9	855	15	US-10-295-027-1185
16	4676	99.9	855	15	US-10-072-012-353
17	4676	99.9	855	15	US-10-072-012-412
18	4676	99.9	855	15	US-10-072-012-419
19	4672	99.8	855	15	US-10-072-012-354
20	4672	99.8	855	15	US-10-072-012-420
21	4672	99.8	855	15	US-10-017-417-132
22	4631	98.9	851	15	US-10-276-774-1798
23	4631	98.9	851	15	US-10-296-115-1143
24	4319	92.3	932	18	US-10-450-763-51801
25	4175.5	89.2	782	14	US-10-097-340-312
26	4175	89.2	762	16	US-10-729-807-1
27	4111	87.8	757	15	US-10-072-012-44
28	3901	83.3	855	9	US-09-900-751-2
29	3901	83.3	855	15	US-10-072-012-355
30	3901	83.3	855	15	US-10-072-012-413
31	3883	83.0	855	15	US-10-072-012-356
32	3883	83.0	855	15	US-10-072-012-414
33	3883	83.0	855	15	US-10-072-012-417
34	3810	81.4	902	15	US-10-333-743-3
35	3810	81.4	902	15	US-10-600-187-10
36	3810	81.4	902	16	US-10-297-987B-11
37	2980	63.7	620	9	US-09-925-301-1193
38	2664	56.9	845	15	US-10-072-012-415
39	1352	28.9	362	18	US-10-450-763-51800
40	1319	28.2	241	10	US-09-776-191-50
41	1319	28.2	241	14	US-10-099-700A-4
42	1319	28.2	241	14	US-10-092-004A-2
43	1319	28.2	241	14	US-10-190-030B-4
44	1319	28.2	241	14	US-10-302-840A-4
45	1319	28.2	241	14	US-10-267-219-4

ALIGNMENTS

RESULT 1

US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien

Query Match 100.0%; Score 4681; DB 10; Length 855;

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGPRWVVLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGPRWVVLAA 60

QY 61 VLIGLLVLLGIGIFLWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGIFLWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180
Db 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180

QY 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180
Db 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180

QY 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240
Db 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240

QY 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240
Db 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240

QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PNINDCTWNIETVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT 420
Db 361 PNINDCTWNIETVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SQQNGKDDCGSDGDEASCPKVNVTCTKTYRCLNGCLSKGNPECDGKEDCSGSD 600
Db 541 SQQNGKDDCGSDGDEASCPKVNVTCTKTYRCLNGCLSKGNPECDGKEDCSGSD 600

QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP 720

QY 721 ABYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINOTTENLL 780
Db 721 ABYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINOTTENLL 780

QY 781 PQOITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP 840
Db 781 PQOITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP 840

QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

```

```

RESULT 2
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication NO. US20030008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON

```

```

; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-099-700A-2

```

```

Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGPRWVVLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGPRWVVLAA 60

QY 61 VLIGLLVLLGIGIFLWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGIFLWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180
Db 121 KDALKLLYSGVPLGYPHYKESAVTAFSEGSVIAYYWSFSPHQLVEEAERVMABERVVM 180

QY 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240
Db 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVORTQNSCSFGLHARGVELMRTTFPGFPPSPYPA 240

QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PNINDCTWNIETVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT 420
Db 361 PNINDCTWNIETVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SQQNGKDDCGSDGDEASCPKVNVTCTKTYRCLNGCLSKGNPECDGKEDCSGSD 600
Db 541 SQQNGKDDCGSDGDEASCPKVNVTCTKTYRCLNGCLSKGNPECDGKEDCSGSD 600

QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP 720

QY 721 ABYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINOTTENLL 780
Db 721 ABYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINOTTENLL 780

QY 781 PQOITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP 840
Db 781 PQOITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP 840

```

```
QY      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

RESULT 3
US-10-190-030B-2
; Sequence 2, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190,030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-190-030B-2

Query Match      100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGS DRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60
Db      1 MGS DRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60

QY      61 VLI GLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db      61 VLI GLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY      121 KDALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180
Db      121 KDALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180

QY      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240
Db      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240

QY      241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300
Db      241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300

QY      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360

QY      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
Db      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420

QY      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGTCIRKELRCDCGWADCTDH 480
Db      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGTCIRKELRCDCGWADCTDH 480

QY      481 SDELINCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQCSCPAQTFRCNSGKCLSK 540
Db      481 SDELINCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEQCSCPAQTFRCNSGKCLSK 540

QY      541 SQQCNKGKDCGQSDSEASCPKVVVTCTKHTVRCNLGLCLSKGNPECDGKEDCSGDBK 600
Db      541 SQQCNKGKDCGQSDSEASCPKVVVTCTKHTVRCNLGLCLSKGNPECDGKEDCSGDBK 600

QY      601 DCDCCGLRSFTQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db      601 DCDCCGLRSFTQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
```

```
QY      661 DRGFYSDPTQWTAFLGLHDQSORSAPGVQERRLKRISHPPFNDFTFDYDIALLEKXP 720
Db      661 DRGFYSDPTQWTAFLGLHDQSORSAPGVQERRLKRISHPPFNDFTFDYDIALLEKXP 720

QY      721 ABYSNVRPILCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db      721 ABYSNVRPILCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780

QY      781 POQITPRMCMCVGLSGGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAORNKPGV 840
Db      781 POQITPRMCMCVGLSGGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAORNKPGV 840

QY      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

RESULT 4
US-10-302-840A-2
; Sequence 2, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENC
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-302-840A-2

Query Match      100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGS DRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60
Db      1 MGS DRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60

QY      61 VLI GLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db      61 VLI GLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY      121 KDALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180
Db      121 KDALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180

QY      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240
Db      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240

QY      241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300
Db      241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300

QY      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360

QY      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420
Db      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFV 420

QY      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGTCIRKELRCDCGWADCTDH 480
```


Db 421 NSNKITVRHSDQSYTDITGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
QY 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSQCPAQTFRCSNGKCLSK 540
Db 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSQCPAQTFRCSNGKCLSK 540
QY 541 SQQCNKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
Db 541 SQQCNKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DDCGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQSORQAPGVQERLKRRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQSORQAPGVQERLKRRIISHPPFNDFTFDYDIALLELEKP 720
QY 721 AEYSSWVRPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AEYSSWVRPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 5
US-10-267-219-2
; Sequence 2, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1521
; CURRENT APPLICATION NUMBER: US/10/267,219
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-267-219-2
Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPKDFGAGLKYSRHEKYNLEEGVEFLPVNNVKKVEKGGPRWVWVLA 60
Db 1 MGSDRARKGGGPKDFGAGLKYSRHEKYNLEEGVEFLPVNNVKKVEKGGPRWVWVLA 60
QY 61 VLIGLLLVLLGLGFLVHLYQYRDVVRQKVFNGVMYRITNFENFVDAYNSNSTEFSVLSK 120
Db 61 VLIGLLLVLLGLGFLVHLYQYRDVVRQKVFNGVMYRITNFENFVDAYNSNSTEFSVLSK 120
QY 121 KDALKLLYGVPLGFPYHKESAVTAFSEGSVIAYYWSFSPHQLVVEAERVAERVVM 180
Db 121 KDALKLLYGVPLGFPYHKESAVTAFSEGSVIAYYWSFSPHQLVVEAERVAERVVM 180
QY 181 LPPRARSLSKSVVTSVVAFTDSTKVQRTQDNSCSFGLHARGVELMRTFTPGFPPSPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFTDSTKVQRTQDNSCSFGLHARGVELMRTFTPGFPPSPYPA 240

QY 241 HARCOWALRGDADSVLSLTFRSDFLASCDERGSDLVTVYNTLSMPHEHALVOLCGTTPS 300
Db 241 HARCOWALRGDADSVLSLTFRSDFLASCDERGSDLVTVYNTLSMPHEHALVOLCGTTPS 300
QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPFRMSSCGGRLRKAQGTFRNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPFRMSSCGGRLRKAQGTFRNSPYPGHY 360
QY 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRHSDQSYTDITGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
Db 421 NSNKITVRHSDQSYTDITGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
QY 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSQCPAQTFRCSNGKCLSK 540
Db 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSQCPAQTFRCSNGKCLSK 540
QY 541 SQQCNKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
Db 541 SQQCNKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DDCGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQSORQAPGVQERLKRRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQSORQAPGVQERLKRRIISHPPFNDFTFDYDIALLELEKP 720
QY 721 AEYSSWVRPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AEYSSWVRPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 6
US-10-112-221A-2
; Sequence 2, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-112-221A-2
Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 841 RLPLFRDWIKENTGV 855

RESULT 8

US-10-147-211A-2

; Sequence 2, Application US/10147211A

; Publication No. US20030285900A1

; GENERAL INFORMATION:

; APPLICANT: Madison, Edward

; APPLICANT: Yeh, Jiunn-Chern

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1

; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON

; FILE REFERENCE: 24745-1516

; CURRENT APPLICATION NUMBER: US/10/147,211A

; PRIOR FILING DATE: 2002-05-14

; PRIOR FILING DATE: 2001-05-14

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-147-211A-2

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVKEHGGGRWVLA 60

Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVKEHGGGRWVLA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

QY 121 KOALKLLYSGVPLGYPYHKEASVAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180

Db 121 KOALKLLYSGVPLGYPYHKEASVAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVTSVVAFTDSKTQVTDNSCSFGLHARGVELMRFTTGGFPDPSPPA 240

Db 181 LPPRARSLSKSFVTSVVAFTDSKTQVTDNSCSFGLHARGVELMRFTTGGFPDPSPPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVNTLSPMEPHALVOLCGTYPPS 300

Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLTINTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPGHY 360

Db 301 YNLTFHSSQNVLLITLTINTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPGHY 360

QY 361 PPNIDCTWNI EVNPNQHVSKFVLLPQVAGTCKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNI EVNPNQHVSKFVLLPQVAGTCKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRPHSQSYTDTGFLAEYLSYSDSDPCPGOFTCTGRCIRKELRCGWDCTDH 480

Db 421 NSNKITVRPHSQSYTDTGFLAEYLSYSDSDPCPGOFTCTGRCIRKELRCGWDCTDH 480

QY 481 SDPLNCSCDAGHFTCKNFKCPKPLFWVCDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540

Db 481 SDPLNCSCDAGHFTCKNFKCPKPLFWVCDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540

QY 541 SQQCKGKDCGDSDEASCPKNNVTCTKHTYRCLNGLCLSKGNPECKGKEDCSGDEK 600

Db 541 SQQCKGKDCGDSDEASCPKNNVTCTKHTYRCLNGLCLSKGNPECKGKEDCSGDEK 600

QY 601 DCDCLGRSFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Db 601 DCDCLGRSFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSPTQWTAFLGLHDQSRGAPGVQERRLKRIISHPFNFDFDYDIALLELEXP 720

Db 661 DRGFRYSPTQWTAFLGLHDQSRGAPGVQERRLKRIISHPFNFDFDYDIALLELEXP 720

QY 721 ABYSWVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCE 780

Db 721 ABYSWVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCE 780

QY 781 POQITPRMWCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSWGDCGCAORNKPGVVT 840

Db 781 POQITPRMWCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSWGDCGCAORNKPGVVT 840

QY 841 RLPLFRDWIKENTGV 855

Db 841 RLPLFRDWIKENTGV 855

RESULT 9

US-10-156-214A-2

; Sequence 2, Application US/10156214A

; Publication No. US20040001801A1

; GENERAL INFORMATION:

; APPLICANT: Edwin L. Madison

; APPLICANT: Joseph Edward Semple

; APPLICANT: George P. Vlaeuk

; APPLICANT: Scott Jeffrey Kemp

; APPLICANT: Mallareddy Komandla

; APPLICANT: Daniel Vanna Siev

; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: 24745-1611

; CURRENT APPLICATION NUMBER: US/10/156,214A

; CURRENT FILING DATE: 2002-05-23

; NUMBER OF SEQ ID NOS: 611

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-156-214A-2

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVKEHGGGRWVLA 60

Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVKEHGGGRWVLA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

QY 121 KOALKLLYSGVPLGYPYHKEASVAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180

Db 121 KOALKLLYSGVPLGYPYHKEASVAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVTSVVAFTDSKTQVTDNSCSFGLHARGVELMRFTTGGFPDPSPPA 240

Db 181 LPPRARSLSKSFVTSVVAFTDSKTQVTDNSCSFGLHARGVELMRFTTGGFPDPSPPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVNTLSPMEPHALVOLCGTYPPS 300

Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLTINTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPGHY 360

Db 301 YNLTFHSSQNVLLITLTINTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPGHY 360

QY 361 PPNIDCTWNI EVNPNQHVSKFVLLPQVAGTCKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNI EVNPNQHVSKFVLLPQVAGTCKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
Db 421 NSNKITVRHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SQQCNKGKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600
Db 541 SQQCNKGKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600
QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDOSORSAPGVQERLKEII SHPPNDFTFYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDOSORSAPGVQERLKEII SHPPNDFTFYDIALLELEKP 720
QY 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCCENLL 780
Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDSCGPLSSVEADGRIFQAGVWSWGDCAQRNKGVT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDSCGPLSSVEADGRIFQAGVWSWGDCAQRNKGVT 840
QY 841 RLPLFRDWTIKENTGV 855
Db 841 RLPLFRDWTIKENTGV 855

RESULT 10

US-10-072-012-352

; Sequence 352, Application US/10072012

; Publication No. US20040033493A1

; GENERAL INFORMATION:

; APPLICANT: Tchernev, Velizar

; APPLICANT: Spytek, Kimberly

; APPLICANT: Zernusen, Bryan

; APPLICANT: Patturajan, Meera

; APPLICANT: Shimkets, Richard

; APPLICANT: Li, Li

; APPLICANT: Gangolli, Baha

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Anderson, David W.

; APPLICANT: Rastelli, Luca

; APPLICANT: Miller, Charles E.

; APPLICANT: Gerlach, Valerie

; APPLICANT: Taupier, Jr. Raymond J.

; APPLICANT: Gusev, Vladimir Y.

; APPLICANT: Colman, Steven D.

; APPLICANT: Wolenc, Adam R.

; APPLICANT: Pena, Carol E. A.

; APPLICANT: Furtak, Katarzyna

; APPLICANT: Grosse, William M.

; APPLICANT: Alsobrook II, John P.

; APPLICANT: Lepley, Denise M.

; APPLICANT: Rieger, Daniel K.

; APPLICANT: Burgess, Catherine E.

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-258

; CURRENT APPLICATION NUMBER: US/10/072,012

; CURRENT FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: 60/265,102

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: 60/265,514

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,412

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 352
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-352

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGPKDGFAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPRWVLLAA 60

Db 1 MGSDRARKGGGPKDGFAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPRWVLLAA 60

QY 61 VLIGLLLVLLGIGFLVWHLQYRDVVRQVKNFYNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVWHLQYRDVVRQVKNFYNGYMRITNENFVDAYENSNSTEFVSLASKV 120

QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYTWSSEFSIPOHLVEAERVMASERVVM 180

Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYTWSSEFSIPOHLVEAERVMASERVVM 180

QY 181 LPPRARSLSFWVTSVVAFPDTSKTQVORTQNSCSFGLHARGVELMRFTTPOGFPDPSVPA 240

Db 181 LPPRARSLSFWVTSVVAFPDTSKTQVORTQNSCSFGLHARGVELMRFTTPOGFPDPSVPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTTPPS 300

Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSMPHEHALVOLCGTTPPS 300

QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQOLPRMSSCGRLRKAQGTNSPYPGHY 360

Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQOLPRMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PPNIDCTWNIIEVPNNQHVVSFKFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNIIEVPNNQHVVSFKFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480

Db 421 NSNKITVRHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480

QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCNSGKCLSK 540

Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAQTFRCNSGKCLSK 540

QY 541 SQQCNKGKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600

Db 541 SQQCNKGKDDCGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600

QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDOSORSAPGVQERLKEII SHPPNDFTFYDIALLELEKP 720

Db 661 DRGFRYSDPTQWTAFLGLHDOSORSAPGVQERLKEII SHPPNDFTFYDIALLELEKP 720

QY 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCCENLL 780
Db AYSVMVRPCLPDPASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCCENLL 780
QY 781 PQQITPRMCMVGLSGVDSGCGSLSSVEADGRIFQAGVVSWGDCGAQRNKPQYVT 840
Db PQQITPRMCMVGLSGVDSGCGSLSSVEADGRIFQAGVVSWGDCGAQRNKPQYVT 840
QY 841 RLPLFRDWIKENTGV 855
Db RLPLFRDWIKENTGV 855

RESULT 11
US-10-072-012-411
; Sequence 411, Application US/10072012
; Publication No. US20040053493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 411
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-411

Query Match 100.0%; Score 4681; DB 15; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEBGFPLVNNVKKVEKGGPRWVVLAA 60
Db 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEBGFPLVNNVKKVEKGGPRWVVLAA 60
QY 61 VLIIGLLVLLGIGFLVHLOYRDVROKVFNGYWRITNENFVDAYENSNSSTFVSLASKV 120
Db 61 VLIIGLLVLLGIGFLVHLOYRDVROKVFNGYWRITNENFVDAYENSNSSTFVSLASKV 120
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
QY 181 LPPRARSLSFVVTSSVAFPTDSKTQVTONSCSFGHLHARGVELMRFTTGGFPDPSYPA 240
Db 181 LPPRARSLSFVVTSSVAFPTDSKTQVTONSCSFGHLHARGVELMRFTTGGFPDPSYPA 240
QY 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPFRMSSCGGRLRKAQGTNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPFRMSSCGGRLRKAQGTNSPYYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKSFKFYLLLEPGVPAGTCKPKYVEINGEKYGGERSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVKSFKFYLLLEPGVPAGTCKPKYVEINGEKYGGERSQFVVT 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPGQFTCRGTCRIRKELRCDGWADCTDH 480
Db 421 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPGQFTCRGTCRIRKELRCDGWADCTDH 480
QY 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSPAQTRFCSNGKCLSK 540
Db 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSPAQTRFCSNGKCLSK 540
QY 541 SOQCNKGKDCGDSDEASCPKVVVVTCTKTYRCINGLCLSKGNPECDGKEDCSGSDK 600
Db 541 SOQCNKGKDCGDSDEASCPKVVVVTCTKTYRCINGLCLSKGNPECDGKEDCSGSDK 600
QY 601 DCDGLRSFTRQARVVGTDADGEWPNQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
Db 601 DCDGLRSFTRQARVVGTDADGEWPNQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQERRLKRITISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQERRLKRITISHPFFNDFTFDYDIALLELEKP 720
QY 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCCENLL 780
Db 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTTCCENLL 780
QY 781 PQQITPRMCMVGLSGVDSGCGSLSSVEADGRIFQAGVVSWGDCGAQRNKPQYVT 840
Db 781 PQQITPRMCMVGLSGVDSGCGSLSSVEADGRIFQAGVVSWGDCGAQRNKPQYVT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 12
US-10-072-012-418
; Sequence 418, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly

```
APPLICANT: Zerhusen, Bryan
APPLICANT: Patturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Baha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier, Jr. Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A.
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
FILE REFERENCE: 21402-258
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 418
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-418

Query Match 100.0%; Score 4681; DB 15; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSRRKGGGKDPFGAGLKYNSRHEKVGLEEGVEFLPNNVKKVEKHGPGRWVLA 60
Db 1 MGSRRKGGGKDPFGAGLKYNSRHEKVGLEEGVEFLPNNVKKVEKHGPGRWVLA 60

Qy 61 VLIGLLLVLLGIGFLVHQLQYRDVRQVFNQYMRITNFNFDAVYNSNSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHQLQYRDVRQVFNQYMRITNFNFDAVYNSNSTEFVSLASKV 120

Qy 121 KDALKLLYSVPRLGPGYHKESAVTAFSGSVIAYVWSEFSIPQHLVEAEARVMAEERVVM 180
Db 121 KDALKLLYSVPRLGPGYHKESAVTAFSGSVIAYVWSEFSIPQHLVEAEARVMAEERVVM 180

Qy 181 LPPRARSLSFVTSVVAFPDTSKTQVQTONSCSFGHARGVELMRPTTTPGPDSPYPA 240
Db 181 LPPRARSLSFVTSVVAFPDTSKTQVQTONSCSFGHARGVELMRPTTTPGPDSPYPA 240
```

```
Qy 241 HARCOWALRGDADSVLSLTFERSFDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPBS 300
Db 241 HARCOWALRGDADSVLSLTFERSFDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPBS 300

Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFBEATFFQLPRMSSCGGRLRKAQGTFFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFBEATFFQLPRMSSCGGRLRKAQGTFFNSPYYPGHY 360

Qy 361 PPNIDCTWNIIEVPNNQHVKSFKPFYLLPDPGVPAGTCDPKDYVEINGEKYGGRSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVKSFKPFYLLPDPGVPAGTCDPKDYVEINGEKYGGRSQFVVT 420

Qy 421 NSNKITVRPHSDQSYTDGFLAEVLSYDSSDPCQGFCTCRGTCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDGFLAEVLSYDSSDPCQGFCTCRGTCIRKELRCDGWADCTDH 480

Qy 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540

Qy 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCNLGLCLSKGNPECDGKEDCSGDSDEK 600
Db 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCNLGLCLSKGNPECDGKEDCSGDSDEK 600

Qy 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYD 660
Db 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYD 660

Qy 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQRRRLKRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQRRRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTALILQKGEIRVINOTTENLL 780
Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWTGHTQYGGTALILQKGEIRVINOTTENLL 780

Qy 781 PQQITPRMCMVGLSGGVDSQSGDSGGPLSSVEADGRIFQAGVYVSWGDCQQRNKPQVYT 840
Db 781 PQQITPRMCMVGLSGGVDSQSGDSGGPLSSVEADGRIFQAGVYVSWGDCQQRNKPQVYT 840

Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
```

RESULT 13

```
US-10-600-187-2
; Sequence 2, Application US/10600187
; Publication No. US20040086910A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/10/600,187
; PRIOR APPLICATION NUMBER: US/09/654,600A
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-10-600-187-2
```

Query Match	100.0%;	Score 4681;	DB 15;	Length 855;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 855;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGSDRARKGGGPKDPFAGLKYNSRHEKVNGLKEGVFLPVNNVKYKKGKGGPRWVLA	60	
Db	1	MGSDRARKGGGPKDPFAGLKYNSRHEKVNGLKEGVFLPVNNVKYKKGKGGPRWVLA	60	
Qy	61	VLIGLLLVLLGIGFLVHLOQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120	
Db	61	VLIGLLLVLLGIGFLVHLOQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120	
Qy	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEERVAERVVM	180	
Db	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEERVAERVVM	180	
Qy	181	LPPRARSLKSFVVTSSVAFPTDSKTQVORTQDNSCSFGLHARGVELMRFTTTPGPDSPYPA	240	
Db	181	LPPRARSLKSFVVTSSVAFPTDSKTQVORTQDNSCSFGLHARGVELMRFTTTPGPDSPYPA	240	
Qy	241	HARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPMEPHALVOLCGTYPPS	300	
Db	241	HARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPMEPHALVOLCGTYPPS	300	
Qy	301	YNLTFFHSSQNVLLITLITNTERRHPGFEATPFQOLPRMSSCGRLRKAOGTFNSPYPGHY	360	
Db	301	YNLTFFHSSQNVLLITLITNTERRHPGFEATPFQOLPRMSSCGRLRKAOGTFNSPYPGHY	360	
Qy	361	PNIDCTWNIEVNNQHVSKPKFYLLFEPGVAGTCKPDYVEINGEKYCGERSQFVVT	420	
Db	361	PNIDCTWNIEVNNQHVSKPKFYLLFEPGVAGTCKPDYVEINGEKYCGERSQFVVT	420	
Qy	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH	480	
Db	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH	480	
Qy	481	SDELNCSADAGHFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK	540	
Db	481	SDELNCSADAGHFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK	540	
Qy	541	SQOCNGKDDCGDSDEASCPKVNVTCTKTYRCLNGLCLSKGNPECDGKEDCSGDSDEK	600	
Db	541	SQOCNGKDDCGDSDEASCPKVNVTCTKTYRCLNGLCLSKGNPECDGKEDCSGDSDEK	600	
Qy	601	DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNMLVSAHACYID	660	
Db	601	DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNMLVSAHACYID	660	
Qy	661	DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTFDYDIALLELEKP	720	
Db	661	DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTFDYDIALLELEKP	720	
Qy	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINOTTENLL	780	
Db	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINOTTENLL	780	
Qy	781	PQOITPRMCMVGFSLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYVT	840	
Db	781	PQOITPRMCMVGFSLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYVT	840	
Qy	841	RLPLFRDWIKENTGV	855	
Db	841	RLPLFRDWIKENTGV	855	

RESULT 14
US-10-612-466B-2
; Sequence 2, Application US/10612466B
; Publication No. US20050112579A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; APPLICANT: Yeh, Juinn-Chern

Query Match	100.0%;	Score 4681;	DB 17;	Length 855;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 855;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGSDRARKGGGPKDPFAGLKYNSRHEKVNGLKEGVFLPVNNVKYKKGKGGPRWVLA	60	
Db	1	MGSDRARKGGGPKDPFAGLKYNSRHEKVNGLKEGVFLPVNNVKYKKGKGGPRWVLA	60	
Qy	61	VLIGLLLVLLGIGFLVHLOQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120	
Db	61	VLIGLLLVLLGIGFLVHLOQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120	
Qy	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEERVAERVVM	180	
Db	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEERVAERVVM	180	
Qy	181	LPPRARSLKSFVVTSSVAFPTDSKTQVORTQDNSCSFGLHARGVELMRFTTTPGPDSPYPA	240	
Db	181	LPPRARSLKSFVVTSSVAFPTDSKTQVORTQDNSCSFGLHARGVELMRFTTTPGPDSPYPA	240	
Qy	241	HARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPMEPHALVOLCGTYPPS	300	
Db	241	HARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPMEPHALVOLCGTYPPS	300	
Qy	301	YNLTFFHSSQNVLLITLITNTERRHPGFEATPFQOLPRMSSCGRLRKAOGTFNSPYPGHY	360	
Db	301	YNLTFFHSSQNVLLITLITNTERRHPGFEATPFQOLPRMSSCGRLRKAOGTFNSPYPGHY	360	
Qy	361	PNIDCTWNIEVNNQHVSKPKFYLLFEPGVAGTCKPDYVEINGEKYCGERSQFVVT	420	
Db	361	PNIDCTWNIEVNNQHVSKPKFYLLFEPGVAGTCKPDYVEINGEKYCGERSQFVVT	420	
Qy	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH	480	
Db	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH	480	
Qy	481	SDELNCSADAGHFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK	540	
Db	481	SDELNCSADAGHFTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK	540	
Qy	541	SQOCNGKDDCGDSDEASCPKVNVTCTKTYRCLNGLCLSKGNPECDGKEDCSGDSDEK	600	
Db	541	SQOCNGKDDCGDSDEASCPKVNVTCTKTYRCLNGLCLSKGNPECDGKEDCSGDSDEK	600	
Qy	601	DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNMLVSAHACYID	660	
Db	601	DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNMLVSAHACYID	660	
Qy	661	DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTFDYDIALLELEKP	720	
Db	661	DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTFDYDIALLELEKP	720	
Qy	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINOTTENLL	780	
Db	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINOTTENLL	780	
Qy	781	PQOITPRMCMVGFSLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYVT	840	
Db	781	PQOITPRMCMVGFSLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQYVT	840	

Db 781 PQQITPRMCMVGLSGGVDSQQSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 15
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185

Query Match 99.9%; Score 4676; DB 15; Length 855;
Best Local Similarity 99.9%; Pred No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGSRRARKGGGPKDFGAGLKYNSRHEKVNGLGVEFLPVNNVKKVEKHGFGRWVLA 60
Db 1 MGSRRARKGGGPKDFGAGLKYNSRHEKVNGLGVEFLPVNNVKKVEKHGFGRWVLA 60

Qy 61 VLIGLLLVLLGFLVHLQYRDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120
Db 61 VLIGLLLVLLGFLVHLQYRDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120

Qy 121 KDALKLLYSGVPLGLPYHKESAVTAFSGSVIAYYWSFSIPQHLVEAEERVAEERVVM 180
Db 121 KDALKLLYSGVPLGLPYHKESAVTAFSGSVIAYYWSFSIPQHLVEAEERVAEERVVM 180

Qy 181 LPPRARSLSKSFVTSVVAFTDSKTQRTQDQNSCSFGLHARGVELMRFTTGPFPDSPYPA 240

Search completed: September 23, 2005, 13:12:20
Job time : 178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model
Run on: September 23, 2005, 12:48:10 ; Search time 23 Seconds
(without alignments)
3576.754 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSRRKRGKGGPKDFGAGL.....PGVTRPLPRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3883	83.0	855	2 JC7731	membrane-bound arg
2	727	15.5	1034	1 A53663	enteropeptidase (E
3	712.5	15.2	1035	1 A43090	enteropeptidase (E
4	682.5	14.6	1019	1 A56318	enteropeptidase (E
5	663.5	14.2	1113	2 JE0315	low-density lipopr
6	578.5	12.4	1524	2 T30337	polyprotein - Afri
7	574	12.3	699	1 I54763	Ra-reactive factor
8	533.5	11.4	638	1 KQMSPL	plasma kallikrein
9	514.5	11.0	638	1 KQRTPL	plasma kallikrein
10	509.5	10.9	790	1 PLPG	plasma kallikrein
11	506	10.8	613	2 S15468	complement C3b/C4b
12	502	10.7	460	2 B61545	plasma (EC 3.4.21
13	501.5	10.7	786	1 A47547	serine proteinase
14	500	10.7	638	1 KQHUP	plasma kallikrein
15	497	10.6	810	1 PLHU	plasma (EC 3.4.21
16	492.5	10.5	583	2 A29154	complement factor
17	491.5	10.5	812	1 PLMS	plasma (EC 3.4.21
18	491	10.5	416	1 KFBQ	coagulation factor
19	490.5	10.5	812	1 PLBO	plasma (EC 3.4.21
20	490	10.5	417	1 S00845	hepsin (EC 3.4.21
21	486	10.4	810	2 B30848	plasma (EC 3.4.21
22	484.5	10.4	343	1 A57014	proctasin (EC 3.4
23	482	10.3	686	1 A59271	Ra-reactive factor
24	480.5	10.3	416	1 S33777	hepsin (EC 3.4.21
25	480	10.3	625	1 KFHU1	coagulation factor
26	479.5	10.2	461	1 KFHU	coagulation factor
27	478	10.2	810	2 I46260	plasma (EC 3.4.21
28	472.5	10.1	2616	2 A57096	nudel protein prec
29	472	10.1	275	2 S40007	trypsin (EC 3.4.21

RESULT 1
JC7731
membrane-bound arginine-specific serine proteinase precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: JC7731; JC7775
R/Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda J. Biochem. 130, 425-430, 2001
A/Title: Characterization of a membrane-bound arginine-specific serine protease from r
A/Reference number: JC7731; MUID:21421307; PMID:11530019
A/Accession: JC7731
A/Molecule type: mRNA
A/Residues: 1-855 <KIS>
A/Cross-references: UNIPROT:Q9JUI7; DDBJ:AB049189
A/Experimental source: strain Male, 7-week-old
R/Satoml, S.; Yamasaki, Y.; Tezuka, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T. Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
A/Title: A role for membrane-type serine protease (MT-SP1) in intestinal epithelial tu
A/Reference number: JC7775; PMID:11573963
A/Contents: Small intestine
A/Accession: JC7775
A/Molecule type: mRNA
A/Residues: 1-855 <SAT>
A/Cross-references: DDBJ:AB037898
C/Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates i:
lial migration and/or cell loss.
C/Genetics:
A/Gene: mt-sp1
A/Map position: basolateral cell surface
C/Supfamily: membrane-bound arginine-specific serine proteinase
C/keywords: protein digestion
Query Match 83.0%; Score 3883; DB 2; Length 855;
Best Local Similarity 81.1%; Pred. No. 7.6e-246;
Matches 693; Conservative 79; Mismatches 83; Indels 0; Gaps 0;
Qy 1 MGSRRKRGKGGPKDFGAGLKYSNRHKGVLBEQVEFLPVNNVKYKKGPGRWVLA 60
Db 1 MGNRRKAGGGGQDFGAGLKYSNRLENMNGPEGEVFLPVNNAKQVEKGPWWVMA 60
Qy 61 VLIGLLLVLLGIGFLVWHLQYRDVVRQKVFNGYMRITNENFDAYENSTFEFVLSKV 120
Db 61 VVPSFLLLSLMAGLLVWHLFYHVRVIRKQVFNGLRITNENFDAYENSTFEFVLSKV 120
Qy 121 KDALKLLYSGVPLPGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVBEARVMAEERVM 180
Db 121 KEALKLWYSEVPVLGPYHKKSTVTAASEGSVIAYTWSEFSIPHLBEVDRAVERVVT 180
Qy 181 LPPRRLSKSVVTSVVAFTPTSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYA 240
Db 181 LPPRRLSKSVLTSTVVAFTPTDPRMLQRTQDNCSFGLHARGVTRVTRFTTTPGPNPYPA 240

ALIGNMENTS

Db 794 QEVSKIVGNDSDREGAPWVVALYNGQ-LLCGASLVSRDLVSAACHVYG-----RNLE 848
 Qy 669 PTQWTAFLGLHDQORSAPGVQERLKRISHPPFNDFDYDIALLELEKPAEYSSMVR 728
 Db 849 FSKWALGLLWNTSLTSPQIVTRLDELVINPHYNNRRKSDIAMHLEFKVNTDYIQ 908
 Qy 729 FICLPDASHVFPAGKAIWVTGHTGYGTGALILQKGEIRVINOTTENLLPQ-QITPR 787
 Db 909 FICLPEENQVFPGRICSTAGKVIYQSGPADILQEADVLLSNEKCOQMPENITEN 968
 Qy 788 MNCVGLSGVDSCGDSGGLSSVEADGRIFQAGVWSGDCQARNKPGYVTRPLPRD 847
 Db 969 MNCAGYEGGIDSCGDSGGPLMCLE-NNRWLLAGVTSGYQCALPNRPVYARVPKFE 1027
 Qy 848 WIK 850
 Db 1028 WIQ 1030
 RESULT 3
 A43090
 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
 N:Alternate names: enterokinase
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
 C:Accession: A43090; A48874; A61436
 R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A:Reference number: A43090; MUID:94329561; PMID:8052624
 A:Accession: A43090
 A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1035 <KIT>
 A:Cross-references: UNIPROT:P98072; GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411
 A:Experimental source: small intestine
 R:LaVallie, E.R.; Rehemtulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.;
 J. Biol. Chem. 268, 23311-23317, 1993
 A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of h
 A:Reference number: A48874; MUID:94043122; PMID:8226855
 A:Accession: A48874
 A:Molecule type: mRNA
 A:Residues: 801-1035 <LAV>
 A:Cross-references: GB:L19663; NID:G416131; PIDN:AAAL6035.1; PID:G416132
 A:Note: parts of this sequence, including the amino end of the mature protein, were conf
 R:Light, A.; Janska, H.
 J. Protein Chem. 10, 475-480, 1991
 A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
 A:Reference number: A61436; MUID:92189715; PMID:1799406
 A:Accession: A61436
 A:Molecule type: protein
 A:Residues: 801-807, Y', 809-827 <LIG>
 C:Comment: The mechanism of association with the membrane of the intestinal brush border
 ebrane attachment using a signal-anchor sequence.
 C:Comment: Conversion from membrane-bound to soluble forms may involve further processin
 C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light)
 lfade linked
 C:Function:
 A:Description: cleaves propeptide from trypsinogen to produce active trypsin
 A:Pathway: intestinal digestive hydrolase cascade
 C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep
 C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein
 F:22-38/Domain: transmembrane #status predicted <TM>
 F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
 F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
 F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:358-520/Domain: MAM homology <MAM>
 F:542-647/Domain: Clr/Cls repeat homology <CLR>
 F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
 F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>
 F:801-1030/Domain: trypsin homology <TRY>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binc
 F:788-912,826-842,926-993,957-972,983-1011/diulfide bonds: #status predicted
 F:841,892,987/Active site: His, Asp, Ser #status predicted
 Query Match 15.2%; Score 712.5; DB 1; Length 1035;
 Best Local Similarity 23.3%; Pred. No. 1.2e-38;
 Matches 253; Conservative 149; Mismatches 319; Indels 365; Gaps 45;
 Qy 55 WVLAALVIGLILLVLLGIGFLVWHLQVRDVRQVVF--NGYMR-ITNENRFDAYENST 111
 Db 21 FAVLVFVILVLCAGLIAVSWLSIQGSVKDAAFGSHKRGTLKISGATYNPHLODKLSV 80
 Qy 112 EFVSLASKVDKALKLYSGVFPFLGPHYKESAVTAFSEGSVIAYY-----W-SFESISPOH 164
 Db 81 DFKVLAFDIQOMIDDIQSSNLKNEY-KNSRVLFQENGSIIVIFDILLFDQWVSDQNVKEE 139
 Qy 165 LVEEAERVMAERVVM-----LPPRARSLSKSFVTSV-----V 197
 Db 140 LIOGIEANKSSQLVTFHIDLSNIDITASLENFSTISATTSEK-LTTSIPLATPGNYSI 197
 Qy 198 APPTDSKTQVORTDQNSCSFGLHARGVELM-----RFTTPGF 233
 Db 198 ECPDSSL-----CADALKCIAIDLFCDEGELNCPDGSDEDNKTCATACDGRFLTGS 249
 Qy 234 PDS-----PYPAH-----ARCWALRGDADSVLSLTFRSFD----- 264
 Db 250 SGSFEALHYPKPSNNTSAVCKRIIRVNOGLSIQLNDFYFNTYYADVLNIXEYGMGSSKILR 309
 Qy 265 -----LASCDERGSDLV---TVNTLSMPHEHALVQL----- 293
 Db 310 ASLWSNNPGIIRIFSNQVTFATFIQSDS--SDVIGFKVTVYAFNSKELNNYEKINCFED 367
 Qy 294 --C-----GTYPSPSYNLTF-HS----- 307
 Db 368 GFCEWIDLNDNEWERTQSTFPSTGPTFDHTFGNESGFYSTPTGPGRRRERVGLLT 427
 Qy 308 -----SONVLLITLTNTERRHGFBEATPFQ----- 333
 Db 428 LPLDPTPEQACLSFWYVMYGVNKLINISSDQ---NMEKTIQKEGNYQNNYQGV 484
 Qy 334 -----LPRMSSCGG 342
 Db 485 LNETVEFKVSFYGKQNILSDIALDLSLTGYICNVSVYPEPTLVPTPPPELP--TDCGG 542
 Qy 343 --RLRKAQGFNSPYTPGHPYPPNIDCTWNEVNNQHVKSEKFFYLLERGVGVPAGTCKD 400
 Db 543 PHDLRENTTFTSINFPNSYNOAFCLWNLNAOKGKNIQLHFQBFDELNTA-----D 594
 Qy 401 YVEI-NGEKYCGERSQFV-----VTSNSKITVRFHSDQSYTDGFLAEYLS-- 446
 Db 595 VVEIRDEG---GDSLFLAVYTGPGVPNDVPSTNRMVTLFITDNMLAKQGFKANFTGY 651
 Qy 447 -YDSSDPC-PGQFTCTGTCIRKELRCDCGWADCTDHSDELNC-----SCDAGH--QFTC 496
 Db 652 GLGIPEPKEDNPKQDGEICPLVNLCDGPPHCKDGSDEAHCVRLFNFTDSSGLVQFRI 711
 Qy 497 KNKFCPLFWCVSDVNDGNSDEQSCSPAQTFRCNSGKCLSKSQCNCKDCCGDSDE 556
 Db 712 QS-----IWHV-----ACAEN-----WTQISDDVC-----QLLG---LGTGNS 743
 Qy 557 ASC-----PKNVVVTCTKHTYRCLNGLCLSKGNPECDGKEDCGSDGDEKDCD---CGLR 607
 Db 744 VPTFTSGGPPYVNLNAP-----NGSLILTFSQQC-----LEDLSILLQCNYSKCGK 791
 Qy 608 SFTRQA--RVVGTGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFR 665
 Db 792 LVTOEVSPIKVGSDSREGAWPMWVALYFDQ-QVCCASLVSRDLVLSAAHCYVYG----R 846
 Qy 666 YSDPTQWTAFLGLHDQORSAPGVQERLKRILISHPPFNDFDYDIALLELEKPAEYSS 725
 Db 847 NMEPSKWKAVGLHMAHSLNTSPQIETRLDQIVINPHYNRRKKNNDIAMHLEMKVNYTD 906
 Qy 726 MVRPICLPDASHVFPAGKAIWVTGHTGYGGTGALILQKGEIRVINOTTENLLPQ-QI 784

Db 907 YIPICLPENQYFPFGERICSTAGWALYQGSTADVLQEAADVPLLSNEKCCQOMPEYNI 966
QY 785 TPMMCMVGLSGVDSQCGDGGPLSSVEADGRIFQAGVVSWGDCGAQNRKPGVYTRPL 844
Db 967 TENMVCAGYAGVDSQCGDGGPLMQCB-NRRLWLAGVTSFGYQCALNRRGVARVPR 1025
QY 845 FRDWIK 850
Db 1026 FTEWIK 1031

RESULT 4
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C>Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 09-Jul-2004
C:Accession: A56318; B43090
R:Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A:Reference number: A56318; MUID:95234679; PMID:7718557
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <KIT>
A:Cross-references: UNIPROT:P98073; GB:U09860; NID:G746412; PID:AAC50138.1; PID:G746413
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: B43090
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 749-1019 <KI2>
A:Cross-references: GB:U09860
A:Comment: The mechanism of association with the membrane of the intestinal brush border
oated below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Clis repeat homology; LDL receptor ligand-binding rep
F:1-784/Product: enteropeptidase; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: Clr/Clis repeat homology <Clr>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRQ
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 14.6%; Score 682.5; DB 1; Length 1019;
Best Local Similarity 32.5%; Pred. No. 1.1e-36;
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;

QY 333 QLRMRSSCGG--RLRAQGTFFNSPYYPGHYPNIDCTNIEVNNQHVKVSFKFYLLPEP 390
Db 519 ELP--TDCGPELWEPNTTFSNTPNSYPNLAFVCWILNAQKGNQLQHPQEPDEN- 575
QY 391 GVPAGTCPKDYVEI-NGEK-----YCGERSQFVWTSNKNITVFHSDQSYTDGFL 441

Db 576 -----INDVVEIRDGEADSLLLAVYTGPGVKDVFSTTNRMTVLLITNDVLARGGPK 628
QY 442 AEYL-----YDSSDPCPG-QFTCRTRGCRIRKELRCGWDADCTDHSDELNCS CDAGHOF--- 494
Db 629 ANFTTGHGLPBPCKADHFQCKNGECVPLVNLCDGHLHCEGSDGSDAECV-----RFPNG 683
QY 495 TCKN-----KPKCPILFWVCDSDVNDGSDNDEQGCSCPAOTFRCSNGKCLSKSQOCCNGKDDC 550
Db 684 TTNNGLVRFRIQSIW-----HTACAEN-----WTTQISNDVC-----QLL 719
QY 551 GDGSDASCPKVNVTCTKHTYRCINGLCLSKGNPECDG-----KEDCSDGS-----D 598
Db 720 GLGNSNSKPIFST-----DGOPFVKLNTAPDGHLLITPSQOCLQSLRLQCN 768
QY 599 EKDCDCGLRSFTRQARVVGTTDADEGEWPMQVSLHALQGGHICGASLSPNWLVSAAHCY 658
Db 769 HKSCGKLAADITPKIVGSGNAKEGAPWVVGLY-YGRLLCGASLVSSDWLVSAAHCV 827
QY 659 IDRGFRYSDPTQWTAFLGLHDQORSAPGVQERRLKRIISHPPFNDFTDYDIALLELE 718
Db 828 YG-----RNLEPSKWTAILGLHMKSNLTSPTQVPLIDEIVINPHYNNRRKKNDIAMHLE 883
QY 719 KPAEYSSMVRPICLPDASHVPPAGKAIWVTGWHGTQVGGTGALILQKGEIRVINQTTCE 778
Db 884 FKVNTDYIQICLPEENQVFPGRNCSIAGWTVVYQGTANTILQEAADVPLLSNERCQ 943
QY 779 LLPO-QITPRMCMVGLSGVDSQCGDGGPLSSVEADGRIFQAGVVSWGDCGAQRNPK 837
Db 944 QMPEYNITENMICAGYEEGGIDSCGDSGGPLMQCB-NRWFAGVTSFGYKCALPNRP 1002
QY 838 VYTRPLPLFRDWIK 850
Db 1003 VYARVSRFTEWIK 1015

RESULT 5
JB0315
low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: JB0315
R:Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A:Title: A novel low-density lipoprotein receptor-related protein with type II membran
A:Reference number: JB0315; MUID:98429596; PMID:9756624
A:Accession: JB0315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1113 <TOM>
A:Cross-references: UNIPROT:Q92319; DBJ:AB013874; NID:G3869144; PID:BAJ34371.1; PID:
C:Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor 1
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:784-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:869-1097/Domain: trypsin homology <TRY>

Query Match 14.2%; Score 663.5; DB 2; Length 1113;
Best Local Similarity 32.7%; Pred. No. 2e-35;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

QY 428 RFHSDQSYTDGFLAEVLSYDSDPCPGQFTRGCRIRKELRCGWDADCTDHSDELNCS 487
Db 627 QFPESDQNTCLLP---NEDVECSFHKCRSGRVLGSRRCGQADQDDSDENCG 683
QY 488 CDAGHOFCK-NKPKRPLFWVCDSDVNDGSDNDEQGCSC-CPAQTFRCSNGKCLSKSQOCCN 545
Db 684 CKERALWECFENKCLKHTLICDGFPCDPSMDKNCSCFQCDNELECANECVPRDLWCD 743
QY 546 GKDCGSDGSDASCPKN-----VVTCTKHTYRCLNG-----LC----- 579

Db 744 GWVDCSDSDWECVTLTKXNGNSSLLTVHKSAAEHVHCADGWRETLSLQACKQMLGSP 803
QY 580 -LSKGNPECDGKE-----DCSDGDEK-----C---DCGLRST 610
Db 804 SVTKLPGQGGQWLYPNWENLNGSTQLLVYRHSFSPSRSEISLLSKQDCGRPA 863
QY 611 R-OARVVGTDADGEPWPQVSLHALGQGHICGASLISPNWLSAAHCYIDDRGRFRYSDP 669
Db 864 RMNKRILGRTSRPGRPWQCSLQSEPSGHIICGVLLAKKWLTVAHCF-EGR-----EDA 918
QY 670 TOWTAFGLHDSQORSAPGVQERLKRILSHFPNDFDIDYDIALLELKPARYSMWRP 729
Db 919 DWKVVFGNNLPHPSG-FMOTRFVKTILLHPRYSRAVVDYDISVVLSDDINETSYP 977
QY 730 ICLPDASHVPPAKAKLWTVGWGHTVGGTALILQKGEIRVINQTTCEMLLPQO-ITPRM 788
Db 978 VCLPSPPEEYLEPTYCIYIYNGH--MGNKMPFLQGEVRIIPLOCCQSYFDMKTIITRM 1035
QY 789 MCVGLSGVDSCQDSDGGLSSVEADGRIFQAGVWVGDCGAQR-NKPGVYTRLPFLFRD 847
Db 1036 ICAGYESGTVDSCMGDSGGPLVCERPQGWTLFGLTSWGSVCFKVLGPGVYSNVSYFVG 1095
QY 848 WIK 850
Db 1096 WIE 1098
RESULT 6
T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30337
R:Yang, J.C.; Lindeay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from X
A:Reference number: Z20829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: UNIPROT:091674; EMBL:U01290; NID:G2981640; PID:G2981641; PIDN:AAC247
C:Superfamily: trypsin related polypeptide; trypsin homology
Query Match 12.4%; Score 578.5; DB 2; Length 1524;
Best Local Similarity 26.2%; Pred. No. 1e-29;
Matches 167; Conservative 89; Mismatches 205; Indels 177; Gaps 23;
QY 238 YPAHARQWALRGDADSVLSLTFPSFDL---ASCDERGSGLVTVYNTILS-PMEPHALVQL 293
Db 333 YSINSVCRWMLAVQAKTIEIRFLQDIEDHATC-----TPDYLSFTVNEKMIRKV 383
QY 294 CGTPPSNLTFSHSONVLLITLNTERRHPGFATFPQLP--RMSSCG-GRLRKXQGT 350
Db 384 CGSTIPSLIV--RSNKVTVTFFSDGTFGRGFEIQFLAIPTRKAASACGSAKILKKGM 440
QY 351 FNSPYPCGHPENIDCTWNIENPNHVKVSKFFYLLEPGVPAG-TCPKDYVEI-NG-- 406
Db 441 IYSNPYDPYPRILKTCWIIIEAPENHIVLKPEDF-----NVFYGHCYIDAVEYVDGAE 495
QY 407 -----EKYCGERSQFVWTSNKNITVRPHSPQSYTDTGFLAELIYSDDSPCPQGFCTRT 461
Db 496 EKQIARLCGGVTLPLPISSPENTMLIRPKTDMENSPYGFVKF----- 538
QY 462 GRCIRKELRCDGWADCTDHSBELNCSADAGHQFTCKNK-FCKPLFWVCDSDVNDGDNDSDE 520
Db 539 -----SFVPEKQFSLPV----- 551
QY 521 QGSCPAQTFRCNGKCLSKSQCCNGKDDCGDSDSEASCPKVVVTCVTKHYRCINGLCL 580
Db 552 -----DD-----TPTISML----- 560

QY 581 SKGNPECDKEDCSDGSDKDCDCGLRSFTRO---ARVVGTTDADGEGFWQVSLHALQ 637
Db 561 ---HPRATALD-----VCGMAPMTPKWLPRIVGGEEASPSNWPQVQIFFLRT 606
QY 638 GHIICASLISPNWLSAAHCYIDDRGRFRYSDPTQWTAFLGLHDSQORSAPGVQERLKR 697
Db 607 FH-CBGAIIISQWILTAHC-----IRAAEPSYTVIAGDHNRLNBS-EQIRNIKTI 658
QY 698 LSHFPNDFDIDYDIALLELKPARYSMWRPCLPDASHVPPACKAIWTVGWGHTQYCG 757
Db 659 RIHNNYSETYDNDIALYLEPLDNDLPVCPULPEPEEVLTTPASVCVTVGWGTARDG 718
QY 758 TGAIIQKGEIRVINQTTCE-NLLPQQITPRMVCVGLSG-GVDSQCGDSDGGLSSVEAD 815
Db 719 QPALGLQLQLPILDSIICNTSYSGELTDHMLCAGFPSSKEKDACQDSDGGLVCQNEK 778
QY 816 GRIFQAGVWVGDCGAQRNKPFGVYTRLPFLFRDWIKENT 853
Db 779 EQFSIYGLVSGEGCGRVSKPGVYTKVRLFFTWI-QNT 815
RESULT 7
154763
Ra-reactive factor (EC 3.4.21.-) 1 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
C:Species: Homo sapiens (man)
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: 154763; JN0883
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
Int. Immunol. 6, 665-669, 1994
A:Title: Molecular characterization of a novel serine protease involved in activation
A:Reference number: 154763; MUID:94289349; PMID:8018603
A:Accession: 154763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-699 <SAT>
A:Cross-references: UNIPROT:P48740; GB:D28593; NID:G790963; PIDN:BAA05928.1; PID:G4711
R:Takada, F.; Takayama, Y.; Hatusue, H.; Kawakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
A:Title: A new member of the C1s family of complement proteins found in a bactericidal
A:Reference number: JN0883; MUID:94059062; PMID:8240317
A:Accession: JN0883
A:Molecule type: mRNA
A:Residues: 1-234; 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>
A:Cross-references: DDBJ:DJ17525; NID:G439712; PIDN:BAA04477.1; PID:G439713
A:Experimental source: liver
C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 com
C:Genetics:
A:Gene: GDB:MASP1; GDB:CRARF; CRARF1; PRSS5; MASP
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521
A:Map position: 3q27-3q28
C:Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat hom
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hyd
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>
F:19-135/Domain: C1r/C1s repeat homology <C1R1>
F:143-181/Domain: EGF homology <EGF>
F:185-294/Domain: C1r/C1s repeat homology <C1R2>
F:301-362/Domain: complement factor H repeat homology <FHL>
F:367-432/Domain: complement factor H repeat homology <FHD>
F:449-691/Domain: trypsin homology <TRY>
F:49,178,407/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:73-91,143-157,153-166,168-181,185-212,242-260,301-349,329-362,367-414,397-432,436-57
F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted
F:490,552,646/Active site: His, Ser #status predicted
Query Match 12.3%; Score 574; DB 1; Length 699;
Best Local Similarity 25.3%; Pred. No. 8.4e-30;
Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;
QY 203 SKTVQRTQDNCSCFGLHARGVELMRFPTTPGPPSPYPAHARQWALRGDADSVLSLTFPS 262

Db 15 SKASATVLENNMFG-----QIOSPGYFDS-YPSDSEVTWNITVPDGRFRIKLYFMH 64
QY 263 FOLASCDERGSIDVITYNTLSPMEPHALVOLCG-----TYPPSY-NLTFH 306
Db 65 FNELESSYLCEYDAVKV-----ETEDQVLATFCGRETTDTQTPGQEVVLSFGSPFMSITFR 119
QY 307 SSONVLLIITLITNTERHFGFEATFPQLP----- 335
Db 120 SD-----FENEERFTGDAHYMAVDVDECKEREDBELSCDHCHYCHNYIGGYCSCRF 170
QY 336 -----RMSSCGRLRAQGFNSPYYPGHYPNNIDCTWNIEVNNQHVKSFKF 384
Db 171 GYLITNDNTCTRECSNULFTQRTGVITSPDFNFPKSECELYTELEEGFVNLQFED 230
QY 385 FYLLE--PGVPAGTCKPYVEIN-GEK-----XGERSQFVVTNSNKKITVRHFSQSYTD 437
Db 231 IPDIQDHPEVP--CPYDIKIKVGPVKLPGFCGEKAPETISTQSHSVLILFHSNDAEN 287
QY 438 TQFLABYLSYDSDDPC-----GOFTCRGCRIRKELRCDGWADCTDHSDELNCSDA 490
Db 288 RGWRLSYRA--AGNECEPELQPPVHGKIEPSQAKYFPK-----DQVLVSCDT 331
QY 491 GHOF-----YCKNFKCKPLFW-----VCDVNDGDNDBEQSCCPAQ-----TFR 531
Db 332 GYKVLKDNVEMDTFQIECLKDGWNSKIPTCKIVD-----CRAPGELEHGLITFS 381
QY 532 CSNGKCLSKSQNGKXDCGDSDEASCPKVVNVVTCIKHTYRC-----LNLGLCLSKGNP 585
Db 382 TRNLTYYSEI--KYSQE-----PYRKLNNNTGIYCSAQGVMMKV-LGRSLP 430
QY 586 ECDGKEDCSDGSEKDCDGLRSFTQ--ARVVGTDADGEMWQVLSLHALGQGHICGA 643
Db 431 TCLPV-----CGLPKFSRKLMAFINGRPAQKGTTPMTAMLSHLNGQPPFCGG 477
QY 644 SLISPNWLSAHCY--TD--DRGPRYS--PTOWTAFGLHDQSQSAPGVQERRLK 695
Db 478 SLIGSWMITVAHLQHSJLDGPTLRDSDLSPSDFKIILGKH-WRLRSDNEHGLGVK 536
QY 696 RIISHPFNDFDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQY 755
Db 537 HTTLHPQYDNTFENDVALVELLESPLVNAFVMPICLPGPQ--QEGAMVIVSGWG-KQF 593
QY 756 GOTGALILQGEIRVINQTCNL---LPOQTTPRMCMVGLFSGVDSQCGSGPLSV 812
Db 594 LORFPETLMEIPIVDHSTCKQAVAPLKKVTRDMICAGEKEGKDACAGDSGGPMVTL 653
QY 813 EAD-GRIFOAGVYVSMGDGCAQNRKPGVYTRLPFRDWIKENTGV 855
Db 654 NREGQWYLVGVSMGDDCGKDRYGVYSIIHHNKDWIORVTGV 697

RESULT 8

KQMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemond
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison
A:Reference number: A36557; MUID:91090844; PMID:2264928
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: UNIPROT:P26242; GB:M58588; NID:G200358; PIDN:AAA3393.1; PID:G200359
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:391-621/Domain: trypsin homology <TRY>
F:421-104,477-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,3
F:127,215,308,396,494/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 11.4%; Score 533.5; DB 1; Length 638;
Best Local Similarity 27.1%; Pred. No. 3.3e-27;
Matches 183; Conservative 86; Mismatches 233; Indels 173; Gaps 31;
QY 271 RGSGLVTVNTLSPMEPHALVOLCGTYPSPSYNLTFHSSQNVLITLITNTERHFGF--E 328
Db 32 RGGDLAAIYTP-----DAQYCKMCTFHPRCCLLFSF-----LAVTPPKETNKRFGCFMKE 81
QY 329 ATFFQLPRMSSCG--GELRKAQGFNSPYYPGHYP--NIDCTWNIE-----V 372
Db 82 SITGTLPRIHRTGAISHLSKQCGHQISACHRDYIKGLDMRGSNFNISKTNDNIECQKLC 141
QY 373 PNNQHVK--VSFFKFFYLLE-----PGVPAQT-----C 397
Db 142 TNNFHCQFTVATSAFTRPEYRKCLLKHSASGTPTSKADNLVSGFSLSKALSCSEIGC 201
QY 398 PKD-----YVEINGEKYCGERS---QFVVTNSNKKITVRHFSQSYTDG----FLABY 444
Db 202 PMDIFQHSAPADLVNSQVITPDFAVCRITCITFHFNCPLFFTFYTNWETESQRNVCFKLS 261
QY 445 LSYOSSDPCPQ-----FTCRTR--CIRKELRCDGWADCTDHSDELNCSQDAGH- 492
Db 262 KSGRPSPIPOENAIISGYSLTCTCRTRPEPCHSKI-----YSGVDFFEGLNVTVOGAD 316
QY 493 --QFTCKNFKCKPLFWCDSV-NDGDNDSQSCPAQTRFCNSGKCLSKSQCGNCKDD 549
Db 317 VCQETC-TKTIRCQFFIYSLPQDC---KEEGKC---SURLST----- 353
QY 550 CGDGSDEASCPKVVNVVTCIKHTYRCNLG-----LCLSKGNPECDGKEDCDGSDKDCD 603
Db 354 --DGSP-----TRITYGQSGSYSLALCKLVDPDCTTKIN----- 388
QY 604 CGLSFTFQARVVGTDADGEMWQVLSLHA--LGOQGHICGASLISPNWLSAHCYIDD 661
Db 389 -----ARIVGVTNAGLGEWPMQVLSQVLSQTHLCGSGIIGROWMLVTAACHCF--- 436
QY 662 RGRYSDPTQWTAFLGLHDQSQ--RSAPGVQERRLKRIISHPFNDFDYDIALLELEK 719
Db 437 DGIPY--PDVWRIYGGILSLSEITKTP---SSRIKELIIHQEYKVSSENYDIALIKLOT 491
QY 720 PAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTCNL 779
Db 492 PLNVTEFQKPICLPSKADNTIYTNVWVTGNGYTKEQGETQNILQKATIRLVPNECQKK 551
QY 780 LPQO-ITPRMCMVGLSGVDSQCGDGGPLSSVEADGRIFQAGVVSNGDCCGAQRNKGCV 838
Db 552 YRDVINKQMTICAGYKGGTDAKGDGSGPL-VCKHSGRWQLVGVITSGEGCGRCKDQPGV 610
QY 839 YTRLPLFRDWIKENT 853
Db 611 YTKVSEYMDWILEKT 625

RESULT 9

KORTPL
plasma kallikrein (EC 3.4.21.34) precursor - rat
N:Alternate names: Fletcher factor; kininogenin; serum kallikrein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A39180; A33320; S06851; I53041; S06852
R:Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N
Biochemistry 30, 1628-1635, 1991
A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A:Reference number: A39180; MUID:91129236; PMID:1993180
A:Accession: A39180
A:Molecule type: DNA
A:Residues: 1-638 <BEA>
A:Cross-references: UNIPROT:P14272; GB:J05315
A:Note: the authors translated the codon GAG for residue 81 as Gln
R:Seidah, N.G.; Ladenheim, R.; Miskay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazun
DNA 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: A33320; MUID:90091743; PMID:2598771
A:Accession: A33320
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M30282; NID:G205010; PIDN:AAA41463.1; PID:G205011
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 993, 103-110, 1999
A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A:Reference number: S06851; MUID:90089457; PMID:2597701
A:Accession: S06851
A:Molecule type: protein
A:Residues: 20-45;391-413 <PAQ>
R:Seidah, N.G.; Ladenheim, R.; Miskay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazun
DNA Cell Biol. 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: I53041
A:Accession: I53041
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <RES>
A:Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722
A:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
are linked by one or more disulfide bonds.
C:Genetics:
A:Gene: PK
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-109/Domain: apple repeat <AP1>
F:110-195/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,151-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F:127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:326/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:434,483,578/Active site: His, Asp, Ser #status predicted
Query Match 11.0%; Score 514.5; DB 1; Length 638;
Best Local Similarity 26.5%; Pred. No. 5.8e-26;
Matches 183; Conservative 81; Mismatches 218; Indels 209; Gaps 35;
Qy 271 RGSLLVTVYNTLSPEPHALVOLCTGTPPSYNLTTFHSSQNVLLITLTNTERRHPGF--E 328
Db 32 RGGDLAAYI---TPDAQHC-QKMCFTFPRCLLSP-----LAVPTKETDRRCGFWKE 81
Qy 329 ATPFQLPRMSSCG---GRLRKAQGTFFNSPYTPGHYP-----NIDCTWNIE-----V 372
Db 82 SITGTLPRHRTGATSGHSLKQCGHLSACHQDIYEGLDMRGNSFNISKTDSIECQKLC 141
Qy 373 PNNQHVK---VSEKFFYLE-----PGVPAGTCPKDYVEING--EKYCG----- 411
Db 142 TNNIHCQFFYATKAFHPRPYRKSCLLRSSSGTPTSIKPDVNL-VSGFSLKSCALSSEIG 200
Qy 412 -----ERSQFVTSNSKNTVRHSDQSYTDTGTFLAELVSDSDPCPGFTCTGTGCI 465
Db 201 CPMDFIHFAPADLVNSQVVT-----PDAFVCRV----- 229
Qy 466 KRLRCDCGWADCTDHSDELNGSCDAGHQTCKNFKCKPLF-----WVCDSV-NDGDN 517

Db 230 -----VCTPHP-----NC-----LFFTFYTNWETESQNVCLK 259
Qy 518 SDEQCSCPA-----QTFRCNSGKCLSKSQCKGKDDCG-----DGS 554
Db 260 TSKGRSPFPPIQENAVSGYSLETCRK-----ARPECFKFIYSGVAFEGEELNATFVOGA 315
Qy 555 D--EASCFKVNVTCTKTYRCLNGLCLSKGNPECDGKEDCS-----DGS----- 597
Db 316 DACQETCTK--TIRCQFFYSLLPODCKAEG--C-----KCSLRSLTDSQSPRITYEAQG 366
Qy 598 -----DEKDCDGLRSFTRQARVVGGTDADEGEPWQVSLHA--LQGHICGA 643
Db 367 SSGYSLRLCKVVESSDCTTKI-----NARIVGGTNSLSLGENPQVSLQVLSQNHMCGG 421
Qy 644 SLIPSNVLWSAAHCYIDDRGFYSDDPTOWTAFGLHDQSO--RSAPGVQERLRLKRIISHP 701
Db 422 SIIGQWILTAAHCF---DGIPY--PDVWRIYGGILNLSEITNKT-----FSSIKSLIHHQ 473
Qy 702 PFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGWHGTQYGGTGAL 761
Db 474 KYMSEGSYDIALIKLQTPLYNTEFFQKICLPKSRADTNTIYTNCHVTGWTGTYKGETQN 533
Qy 762 ILQKGEIRVINQTTCCNLLPQO--ITPRMVCVGLSGVDSCQDGGSGPLSSVEADRIFFQ 820
Db 534 ILQKATIPLVNPEECQKKYRDYVITKQMICAGYKEGGIDACKGDSGGPL-VCKHSGRWQL 592
Qy 821 AGVYSGDGGCAQRNKGVTYTLPLFPROWKE 851
Db 593 VGIWSGEGCARKQPGVYTKVAEYIDWILE 623
RESULT 10
PLPG
plasmin (EC 3.4.21.7) precursor - pig (fragment)
N:Alternate names: plasminogen
N:Contains: miniplasminogen
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: S03733; S03737; A25834
R:Schaller, J.; Marti, T.; Roeselelet, S.J.; Kaempfer, U.; Rickli, E.E.
Fibrinolysis 1, 91-102, 1987
A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
A:Reference number: S03733
A:Accession: S03733
A:Molecule type: protein
A:Residues: 1-560 <SCH>
R:Brundage, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M
Eur. J. Biochem. 114, 465-470, 1981
A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of human
A:Reference number: S03735; MUID:81212097; PMID:7238497
A:Accession: S03737
A:Molecule type: protein
A:Residues: 1-57 <BRU>
R:Marti, T.; Schaller, J.; Rickli, E.E.
Eur. J. Biochem. 149, 279-285, 1985
A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.
A:Reference number: A25834; MUID:85203907; PMID:3846533
A:Accession: A25834
A:Molecule type: protein
A:Residues: 450-790 <MAR>
C:Function:
A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen a
A:Pathway: fibrinolysis
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homolo
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine pro
F:1-790/Product: plasminogen #status predicted <PRO>
F:1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>
F:1-77/Domain: activation peptide #status predicted <APT>
F:78-560/Product: plasmin chain A #status predicted <ACH>
F:84-162/Domain: kringle homology <KRI>


```

Query Match      10.6%; Score 497; DB 1; Length 810;
Best Local Similarity 24.6%; Pred No. 1.1e-24;
Matches 224; Conservative 104; Mismatches 294; Indels 2

QY      102 VDAYENSNSTEFVSLASKVKVDALKLLYSGVPFLGPLYHKESAVTAFSEGSV
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      22 LDDYVNTQGLASFVTKQ-----LCAGSIEECAAKCEED--
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      160 -SIPQHLVTEAREVMAEERVVMIPPRARSLKSFVTVTSVVAFTPSK-----
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 62 RAFQHSKEQQCVIMABNRKSIIRMDV---VLFKVKVYLSECKTGNKNGRTGNSKT 118
QY 210 QDNCSGFLHARGVELMRFTT-----PGFPDGPYPAHA---RCO-----WALRGD 251
Db 119 KN-----GITCKWSSSPRPRSPATHSEGLEENYCRNPNDPDGPMWCYTDD 168
QY 252 ADSVLSLTPRSPDLASCDER-----GSDI-VTVYNTLSPE-----PHALVOLCGTY 297
Db 169 PEK---RYDYCDILECEBECMHCSGENYDGKISKTMGLECOAWDSQSPA---HGYI 220
QY 298 P---PSYNTLTHSSQNVLITITWERR-----HPGEATFPOLPRMS-----SCGRL 344
Db 221 PSKFNKNLKNYCRN-----PDRELAPWCFTTDPNKRWELCDIPRCTTTPPPSSGPTY 273
QY 345 RKAQGTFSPPYYPGHYPNIDCT-----WNIEVPNNQHVKSFFFYLLLEPGVPAGT 396
Db 274 QCLKGTE-----NYRGNVAVTVSGHTCOHWSAQTPHTN-----R 309
QY 397 CPKQVVEIN-GEKIC-----GERSQFVVTNSNKKITVRHSDQSYTDTGFLAFLYSYDSSD 451
Db 310 TPENFPCKNLDENYCRNPDGKRAPWCHTTNSQ---VRW-----EYCKIPSCD 353
QY 452 PCP-----LQ-----CO-----FTCRGRCIRKELRCDGWADCTDH 480
Db 354 SSPVSTEQLAPTAPPELTPVQDQYHGDQSYRGTSSTTTGK-----KCQSWSSMTPH 407
QY 481 SDELNCSCDAGHQFTCKNKFKFLFWVCDNVNDGDNDSDEQCSC---PAQTER-CSNG 535
Db 408 R-----HOKTPENYPNAGL-----TMYCRNPADKGPWCFTTDPSPVRWEYCNLK 452
QY 536 KC-----LSKSQOCNGKDDC--GDG-----SDEAS----- 558
Db 453 KCSGTEASVAVPPVLLPNVETPSEEDCMFGNGKGYRGRATTVTGTPODWAQEPHR 512
QY 559 ----CPKNVVTCTKHTYRCLN--GLCLSKGNPEC--DGKEDCSGSEKDC-----DCG 605
Db 513 HSIPTETNPRAGLEKNY--CRNPDG---DVGGPWCYTTNPRKLYDYCDVPQCAAPSFDCG 568
QY 606 LRSFTRQ---ARVVGGTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCIYDDR 662
Db 569 KPQVEPKKCPGRVVGCVAHPHSHWQVSLRTRFGMHFCGGTLISPEWVLTAAHCL--EK 626
QY 663 GFRYSDPTQWTAFLGLHDQSQRSAPGQVRRKRIISHPFFNDFTFDYDIALLELEKPAE 722
Db 627 SPR---PSSYKVLGAH-QEVNLEPHVQIEVSRLEPFRK-----DIALKLSSPAV 676
QY 723 YSSMRPDIPLDASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVINOTTCE--NLL 780
Db 677 ITDKVIPACLPSPNTVAVDRTECFITGNETQ-GTFGAGLLKEAQLPVIENKVCNRYEFL 735
QY 781 PQOITPRMVCFLSGSDSCGSDGGLSSVEADGRIFOAGVVSWGDCGAORNKPGVYT 840
Db 736 NGRVOSTELCAGLAGGTDSCGSDGGLVCFEKDKYILQ-GVTSWGLGCAARNKPGVYV 794
QY 841 RLPLFRDWIK 850
Db 795 RVSREVTWIE 804

Search completed: September 23, 2005, 12:56:06
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:49:05 ; Search time 179 Seconds
(without alignments)
2445.965 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGSDRARKGGGPKDFGAGL.....PGVYTRFLPRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_O3.*

1: uniprot_spot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4676	99.9	855	1 ST14 HUMAN	Q9Y5Y6 homo sapien
2	3901	83.3	855	1 ST14 MOUSE	P56677 mus musculus
3	3883	83.0	855	2 Q9JJI7	Q9JJI7 rattus norv
4	2690	57.5	845	2 Q6GR54	Q6GR54 xenopus lae
5	2664	56.9	845	2 Q9DGR1	Q9DGR1 xenopus lae
6	2653	56.7	845	2 Q6JZQ6	Q6JZQ6 xenopus lae
7	2379	50.8	422	2 Q8WVC1	Q8WVC1 homo sapien
8	1969	42.1	663	2 Q6DEV0	Q6DEV0 xenopus tro
9	1124.5	24.0	799	2 Q6PF94	Q6PF94 mus musculus
10	1124.5	24.0	811	1 TMS6_MOUSE	Q8db10 mus musculus
11	1124	24.0	811	1 TMS6_HUMAN	Q8lu80 homo sapien
12	1122	24.0	802	2 Q6UXD8	Q6uxd8 homo sapien
13	1102	23.5	824	2 Q6ICC2	Q6icc2 homo sapien
14	1011.5	21.6	572	2 Q7RTY8	Q7rty8 homo sapien
15	1011.5	21.6	572	2 Q8BIK6	Q8bik6 mus musculus
16	727	15.5	1034	1 ENTK_PIG	P98074 sus scrofa
17	717.5	15.3	855	2 Q7Z410	Q7z410 homo sapien
18	717.5	15.3	1059	2 Q7Z411	Q7z411 homo sapien
19	712.5	15.2	1035	1 ENTK_BOVIN	P98072 bos taurus
20	692	14.8	1042	1 COR1_HUMAN	Q9Y5G5 homo sapien
21	690.5	14.8	1111	2 Q80YN4	Q80yn4 rattus norv
22	687	14.7	777	2 Q8CAN9	Q8can9 mus musculus
23	682.5	14.6	1019	1 ENTK_HUMAN	P98073 homo sapien
24	676.5	14.5	1069	1 ENTK_MOUSE	P97435 mus musculus
25	663.5	14.2	1113	1 COR1_MOUSE	Q9Z319 mus musculus
26	644.5	13.8	767	2 Q9DGR2	Q9dgr2 xenopus lae
27	644	13.8	722	2 Q6NUF5	Q6nuf5 xenopus lae
28	636.5	13.6	680	2 Q868H7	Q868h7 branchiosto
29	623.5	13.3	680	2 Q868H5	Q868h5 branchiosto
30	617	13.2	581	2 Q9XZM7	Q9xzm7 strongyloce
31	612	13.1	688	2 Q868H6	Q868h6 branchiosto

32	601	12.8	490	2	Q7TN04	Q7tn04 mus musculus
33	600	12.8	490	1	TMS2_MOUSE	Q9jiq8 mus musculus
34	600	12.8	490	2	Q920K3	Q920k3 rattus norv
35	598.5	12.8	490	2	Q6P7D7	Q6p7d7 rattus norv
36	597	12.8	688	2	Q868H4	Q868h4 branchiosto
37	593.5	12.7	701	2	Q9JJS9	Q9jjs9 rattus norv
38	588	12.6	422	1	DESI_HUMAN	Q9ul52 homo sapien
39	588	12.6	423	2	Q6UW31	Q6uw31 homo sapien
40	587.5	12.6	1134	2	Q7RTY7	Q7rty7 homo sapien
41	587	12.5	703	2	Q8CHN8	Q8chn8 rattus norv
42	586.5	12.5	704	1	CRAR_MOUSE	P98064 mus musculus
43	585.5	12.5	868	2	Q9Y1V3	Q9y1v3 polyandroca
44	578.5	12.4	1524	2	Q91674	Q91674 xenopus lae
45	577	12.3	676	2	Q6DUJ6	Q6duj6 cyprinus ca

ALIGNMENTS

RESULT 1

ST14_HUMAN
ID ST14_HUMAN STANDARD; PRT; 855 AA.
AC Q9Y5Y6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrptase) (Membrane-type serine protease 1) (MT-SP1) (Prostamin) (Serine protease TADG-15) (tumor associated differentially-expressed gene-15 protein).
DE Name=ST14; Synonyms=PRSS14, SNC19, TADG15;
GN Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303581; PubMed=10373424; DOI=10.1074/jbc.274.25.18231;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity."
RT J. Biol. Chem. 274:18231-18236(1999).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432178; PubMed=10500122; DOI=10.1073/pnas.96.20.11054;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue."
RT Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RL [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Prostate;
RA Yamaguchi N., Mitsu S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed in human prostate."
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parnley T.H., O'Brien T.J.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood, and Muscle;
RA MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shervenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnersch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SNCL19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425; DOI=10.1074/jbc.274.26.18237;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilysin
RT and a Kunitz-type serine protease inhibitor from human milk";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -!- FUNCTION: Degrades extracellular matrix. Proposed to play a role
CC in breast cancer invasion and metastasis. Exhibits trypsin-like
CC activity as defined by cleavage of synthetic substrates with Arg
CC or Lys as the P1 site.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the peptidase S1 family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF118224; AAD4765.2; -;
CC EMBL; AF133086; AAF00109.1; -;
CC EMBL; AB030036; BAB20376.1; -;
CC EMBL; AF057145; AAG15395.1; -;
CC EMBL; BC005826; AAH05826.2; -;
CC EMBL; BC030532; AAH30532.1; -;
CC EMBL; AF283256; AAG13949.1; -;
CC PDB; 1EAW; X-ray; A/C615-855.
CC MEROPS; S01.302; -;
CC Genew; HGNC:11344; S114.
CC MIM; 606797; -;
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0008236; F:serine-type peptidase activity; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR002172; LDL receptor A.
CC InterPro; IPR009003; Pept Ser_Cys_1.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00431; CUB; 2;
CC Pfam; PF00057; Ldl recept a; 4.
CC Pfam; PF00089; Trypsin; 1;
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00261; LDLRECEPTOR.
CC SMART; SM00042; CUB; 2;
CC SMART; SM00192; LDLA; 4.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS01180; CUB; 2;
CC PROSITE; PS01209; LDLRA_1; 2;
CC PROSITE; PS00668; LDLRA_2; 4.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;
KW Signal-anchor, Transmembrane.
FT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane
FT protein (Potential).
FT EXTRACELL 77 855 Extracellular (Potential).
FT CUB 1.
FT CUB 2.
FT LDL-RECEPTOR 487 524 LDL-receptor class A 1.
FT LDL-RECEPTOR 524 560 LDL-receptor class A 2.
FT LDL-RECEPTOR 560 603 LDL-receptor class A 3.
FT LDL-RECEPTOR 603 615 LDL-receptor class A 4.
FT SERINE PROTEASE 615 854 Serine protease.
FT ACT_SITE 656 656 Charge relay system (By similarity).
FT ACT_SITE 711 711 Charge relay system (By similarity).
FT ACT_SITE 805 805 Charge relay system (By similarity).
FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 485 485 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
FT CONFLICT 381 381 R -> S (in Ref. 4).
FT CONFLICT 674 674 A -> V (in Ref. 3).
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 99.9%; Score 4676; DB 1; Length 855;
Best Local Similarity 99.9%; Pred. No. 1.2e-312;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSRRARKGKGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKGKGGPRWVLA 60
DB 1 MGSRRARKGKGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKGKGGPRWVLA 60

QY 61 VLIGLLVLLGIGFLVWHLQYRVRQKVFNGYVRIITNENFVDAYNSNSTEFVSLASKV 120
DB 61 VLIGLLVLLGIGFLVWHLQYRVRQKVFNGYVRIITNENFVDAYNSNSTEFVSLASKV 120

QY 121 KDALKLYSGVFFLGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVEBAERVAERVVM 180
DB 121 KDALKLYSGVFFLGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVEBAERVAERVVM 180

QY 181 LPPRARSLSKPSVTSVVAFTDSKTQVORTQNSCSFGLHARGVELMRFTTGFDPSPYA 240
DB 181 LPPRARSLSKPSVTSVVAFTDSKTQVORTQNSCSFGLHARGVELMRFTTGFDPSPYA 240

QY 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVYNTLSMPHEHALVOLCTYPPS 300
DB 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVYNTLSMPHEHALVOLCTYPPS 300

QY 301 YNLTFHSSQNVLLITLTITNERRHPGFATPFQLPFRMSSCGRLRKAQGTFSNPPYGHY 360
DB 301 YNLTFHSSQNVLLITLTITNERRHPGFATPFQLPFRMSSCGRLRKAQGTFSNPPYGHY 360

QY 361 PPNIIDCTWNIETVNNQHVSKFKFYLLEPGVPAGTCKDYVEINGEKYQGERQFVYTS 420
DB 361 PPNIIDCTWNIETVNNQHVSKFKFYLLEPGVPAGTCKDYVEINGEKYQGERQFVYTS 420

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELRCDGWDCTDH 480
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYSSDPGQFTCTGRCIRKELRCDGWDCTDH 480

QY 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGDNSEDEGSCPAQTRCNSNGKLSK 540
DB 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGDNSEDEGSCPAQTRCNSNGKLSK 540

QY 541 SQQCNKGKDCGDSDEASCPKVNVTCTKTYRCINGLCLSKGNPECDGKDCSDGSEK 600
DB 541 SQQCNKGKDCGDSDEASCPKVNVTCTKTYRCINGLCLSKGNPECDGKDCSDGSEK 600

QY 601 DCDGLRSFTRQARVVGTDADGEWQVSLHALGQGHICGASLISPNMLVSAHCYID 660
DB 601 DCDGLRSFTRQARVVGTDADGEWQVSLHALGQGHICGASLISPNMLVSAHCYID 660

QY 661 DRGFRYSPTQWTAFLGLHDQSQRAPGVQERLKRILSHPPFNFTFDYDIALLELEKP 720
 Db 661 DRGFRYSPTQWTAFLGLHDQSQRAPGVQERLKRILSHPPFNFTFDYDIALLELEKP 720
 QY 721 AYSWSSWRPILCPDASHVFPAGKATVWTGHTQVGGTGCALILQGEIRVINQTCENLL 780
 Db 721 AYSWSSWRPILCPDASHVFPAGKATVWTGHTQVGGTGCALILQGEIRVINQTCENLL 780
 QY 781 PQOITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQARNKPGVYT 840
 Db 781 PQOITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQARNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 2

ST14 MOUSE
 ID ST14_MOUSE STANDARD; PRT; 855 AA.
 AC P56677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Suppressor of tumorigenicity 14 (SC 3.4.21.-) (Epithin).
 GN Name=St14; Synonyms=Pr8814;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;
 RX MEDLINE=9216440; PubMed=1019918; DOI=10.1007/s002510050515;
 RA Kim M.G., Chen C., Lyl M.S., Cho E.G., Park D., Kozak C.,
 RA Schwartz R.H.;
 RT "Cloning and chromosomal mapping of a gene isolated from thymic
 RT stromal cells encoding a new mouse type II membrane serine protease,
 RT epithin, containing four LDL receptor modules and two CUB domains.";
 RL Immunogenetics 49:420-428(1999).
 RN [2]
 RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Colling P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J.J., Helton E., Young A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Ketting A.C., Shevchenko Y., Bouffard G.G.,
 RA Blackley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,
 CC and thymus. Not expressed in skeletal muscle, liver, heart, testis

CC and brain.
 CC -1- SIMILARITY: Belongs to the peptidase S1 family.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF042822; AAD02230.3; -;
 CC EMBL; BC005496; AAH05496.1; -;
 CC HSSP; P00760; LEZX.
 CC MEROPS; S01.302; -;
 CC MGD; MGI:1338881; St14.
 CC GO; GO:0005576; C:extracellular; IDA.
 CC GO; GO:0019897; C:extrinsic to plasma membrane; IDA.
 CC GO; GO:0008236; F:serine-type peptidase activity; IDA.
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC InterPro; IPR009003; Pept_Ser_Cys.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00057; Ldl_recept_a; 4.
 CC Pfam; PF00089; Trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC SMART; SM00042; CUB; 2.
 CC SMART; SM00192; LDLA; 4.
 CC SMART; SM00020; Tryp_SPC; 1.
 CC PROSITE; PS01180; CUB; 2.
 CC PROSITE; PS01209; LDLRA_1; 2.
 CC PROSITE; PS00068; LDLRA_2; 4.
 CC PROSITE; PS02040; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
 KW Transmembrane.
 FT DOMAIN 1 55 Cytoplasmic (Potential).
 FT TRANSMEM 56 76 Signal-anchor for type II membrane
 FT protein (Potential).
 FT DOMAIN 77 855 Extracellular (Potential).
 FT DOMAIN 214 331 CUB 1.
 FT DOMAIN 340 444 CUB 2.
 FT DOMAIN 451 488 LDL-receptor class A 1.
 FT DOMAIN 489 522 LDL-receptor class A 2.
 FT DOMAIN 523 561 LDL-receptor class A 3.
 FT DOMAIN 565 604 LDL-receptor class A 4.
 FT DOMAIN 615 854 Serine protease.
 FT ACT_SITE 656 656 Charge relay system (By similarity).
 FT ACT_SITE 711 711 Charge relay system (By similarity).
 FT ACT_SITE 805 805 Charge relay system (By similarity).
 FT CARBOHYD 107 107 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 365 365 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 421 421 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 489 489 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;
 Query Match 83.3%; Score 3901; DB 1; Length 855;
 Best Local Similarity 81.8%; Pred. No. 2,1e-259;
 Matches 699; Conservative 73; Mismatches 83; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGPKDFGAGLKYNSRHEKYNGLSEGVFLPNNVKKVKGHGPGRWVLA 60
 Db 1 MGSNRGRKGGGQDFGAGLKYNSRHEKYNGLSEGVFLPNNVKKVKGHGPGRWVLA 60
 QY 61 VLITGLLLVLLGIGFLVWHLQYRDVYQKVFNGYWRITNENFVDAYENSNSFVSLASKV 120

QY 491 GHQFTCKN-KFCXPLFWVCDVNDGNSDEQSCCPAQTFRCSNGKCLSKSQCCNGKDD 549
DR PROSITE; PS01209; LDLRA 1; 2.
Db 479 -LQFRCVNSKLCXPSYFICDGVNDGSDSDDELACKCPNNTFKCGNGKCIIPDSQKCDRVN 537
QY 550 CGDGSDEACPKVNVVCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK--DCDCGLR 607
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
Db 538 CGDGSDEACDQVLTACTEYTKCKNNOCITKKNPECDGENDCSGDSNAKCNCGKR 597
QY 608 SPTQARVVGTTDADGEWPMQVSHALGQGHICGASLISPNWLVSAAHCYIDDRGFYS 667
Db 598 PFTKKSRIYGVGNADTGEFPWQVSLHAKGNKHTCGASLVSPTMLISAACHFQDDHOMRYS 657
QY 668 DPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFNDFTDYDIALLELEKPAEYSNV 727
Db 658 DASLWTAYLGLHDQALNTKDVVERIKRIMAHIGFNDNTYDNDIAVLELEKPEVETDFI 717
QY 728 RPICLPDASHVFPAGKAIWVTGWGHTQYGGTGMALILQKEIRVINQTTCCENLLPQOITPR 787
Db 718 QPVCIPESTHDFPVGKPIWVTGWGALKKEGGAAVILQKAEIRIINQTECNKLLDGLTPR 777
QY 788 MNCVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVVSWMGDGCAORNKPGVYVTRLPDR 847
Db 778 MLCAGFVSGGIDACQDSSGGLSSVELNNKVLAGVVSWMGEGCARRNKEGVYTKVSMRDR 837
QY 848 WIKENTGV 855
Db 838 WSKDTGL 845

RESULT 5
Q9DGR1 PRELIMINARY; PRT; 845 AA.
ID Q9DGR1
AC Q9DGR1
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Homolog of human MT-SPI.
GN Names=MT-SPI;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
RP MEDLINE-20363741; PubMed-10903452; DOI=10.1016/S0378-1119(00)00225-0;
RX Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis."
RL Gene 252:209-216(2000).
CC 1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AB038498; BAB03218.1; -;
DR HSSP; Q9Y516; IEAW.
DR MEROPS; S01.050; -;
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLa; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01209; LDLRA 1; 2.
DR PROSITE; PS00068; LDLRA 2; 4.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;
Query Match 56.9%; Score 2664; DB 2; Length 845;
Best Local Similarity 55.3%; Pred. No. 2,1e-174;
Matches 469; Conservative 154; Mismatches 215; Indels 10; Gaps 5;
QY 14 KDFAGLKYNSRHEKNGLEGEVFLPVNNVKKVKEHGPRVWVLAALVIGLLLVLLGTG 73
Db 2 KDSMMKYNRRPQSLNGFEQVEFLPATNSKKVEKTPKKKLAIFGLVIGALLSLTIG 61
QY 74 FLVHLQVDRVYQVFNQYMRITNENFVDAYENSSTEFVSLASKVDAKLLVSGVFP 133
Db 62 LLVHFAIRNKPVKLYGTITANTPFI DAYENSTTAEFSDLSAKVIDTQTVYNGKD 121
QY 134 LGPYHKEAVTAFSEG---SVIAYYWSFSPQHLVVEAERVMAEERVVMPLPPARSLS 190
Db 122 IAPYLQKCSISAPSEGGNNVIGYWSFSPAPREAAFEKAISELKLPVNPQR--T 178
QY 191 FVVTSVVAFPTDSKTVQRTQNSCFGHLHARGVELMRFTTFCGPDSPYPAHARCOWALRG 250
Db 179 FALDSLVAIPTDQIARVFKNSSCAYFLHSSNGVVAKEFSPGFPDPSYPRNARCLWTLRA 238
QY 251 DADSVLSLTPRSFDLASCDEGSDLVTVYNTLSMPHEPALVOLCGTYPSPYNLTTHSSON 310
Db 239 DAGRIIHLHFKTQMEKCKPNGDFFVMVYDLSPIEPRAQLRLCGIYPPSYNLTTFSSN 298
QY 311 VLLITLITNERRHPGEATFFQLPRMSSCGRLRKAQGTFSNPPYGHYPNIDCTWNI 370
Db 299 VMLVTIVTDNVGKPGFLAEFKQLPKTSLCGLLRDASGFTSPFPFAHYPPSTESINDI 358
QY 371 EVPNQHVKSFKFFYLLEPGVPAGTCKPVVEINGEKYCGERSOFVVTSSNKTIVRPH 430
Db 359 QVPDNKFVKVRFNMFYLAEPGPVPTKTKDFEIKGQKYCGEKEFFVVSNNSSKMSVRV 418
QY 431 SDQSTDTGFLAEVLSYDSDPCGQFCTGRGCIKRLKRCGDGADCTDHSDELNCSDA 490
Db 419 SDQSTDTGFLAEVLSYDSDPCGQFCTGRGCIKRLKRCGDGADCTDHSDELNCSDA 478
QY 491 GHQFTCKN-KFCXPLFWVCDVNDGNSDEQSCCPAQTFRCSNGKCLSKSQCCNGKDD 549
Db 479 -LQFRCVNSKLCXPSYFICDGVNDGSDSDDELACKCPNNTFKCGNGKCIIPDSQKCDRVN 537
QY 550 CGDGSDEACPKVNVVCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK--DCDCGLR 607
Db 538 CGDGSDEACDQVLTACTEYTKCKNNOCITKKNPECDGENDCSGDSNAKCNCGKR 597
QY 608 SPTQARVVGTTDADGEWPMQVSHALGQGHICGASLISPNWLVSAAHCYIDDRGFYS 667
Db 598 PFTKKSRIYGVGNADTGEFPWQVSLHAKGNKHTCGASLVSPTMLISAACHFQDDHOMRYS 657
QY 668 DPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFNDFTDYDIALLELEKPAEYSNV 727
Db 658 DASLWTAYLGLHDQALNTKDVVERIKRIMAHIGFNDNTYDNDIAVLELEKPEVETDFI 717
QY 728 RPICLPDASHVFPAGKAIWVTGWGHTQYGGTGMALILQKEIRVINQTTCCENLLPQOITPR 787
Db 718 QPVCIPESTHDFPVGKPIWVTGWGALKKEGGAAVILQKAEIRIINQTECNKLLDGLTPR 777
QY 788 MNCVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVVSWMGDGCAORNKPGVYVTRLPDR 847
Db 778 MLCAGFVSGGIDACQDSSGGLSSVELNNKVLAGVVSWMGEGCARRNKEGVYTKVSMRDR 837
QY 848 WIKENTGV 855
Db 838 WSKDTGL 845

Q63206	PRELIMINARY;	PRT;	845 AA.	
AC	Q63206;			
DT	25-OCT-2004 (TrEMBLrel. 28, Created)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Hypothetical protein.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RX	MEDLINE=223411132; PubMed=12454917; DOI=10.1002/dvdy.10174;			
RA	Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,			
RA	Richardson P.,			
RT	"Genetic and genomic tools for Xenopus research: The NIH Xenopus			
RL	initiative."			
RL	Dev. Dyn. 225:384-391 (2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RX	PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny P.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Trinchieri J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Goughwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smaluk D.E., Scherch A., Schein J.E.,			
RA	Jones S.J., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RL	Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC082854; AA082854.1; -			
KW	Hypothetical protein.			
SQ	SEQUENCE 845 AA; 93710 MW; 6B15E0E7BAA91A9 CRC64;			
Qy	Query Match	56.7%;	Score 2653;	DB 2; Length 845;
Db	Best Local Similarity	54.6%;	Pred. No. 1.2e-173;	
Qy	Matches 463; Conservative 153; Mismatches 22;		Indels 10; Gaps 5;	
Db	14 KDFGGLKNSRHEKVGLEGEVPELPVNNVKKVKGHPGRVWVLAVLGLLVLIG 73			
Qy	2 KDSIMKYNKPSLNGFEFVEFLPATNSKKVEKTPKKLAIFGVVIGALISLTIG 61			
Db	74 FLVHQLQYRVKQVFNQYRNITNENFVAYNSNTEFVSLAKVDALKLYSGVFP 133			
Qy	62 LLVWHFAYRNAPTHKLYGTITANTQFIEAYNSTTEFADLAKVINTLQTVNGKD 121			
Db	134 LQPYKHSVAFTAREG----SVIAYWSEFSIPQHLVEAEARVMAERVWMLPPRARSILKS 190			
Qy	122 IAPYLQCCISARSEGNNVIGYWSFVPAFREAAFEKAISEFKLPSPVDPRR---T 178			
Db	191 FVTSVVAFPDTSKTVQRTQNSCSFGLHARGVELMRTTPGPDSPYAHARQWALRG 250			
Qy				

RESULT 7
Q8WVC1
ID Q8WVC1
AC Q8WVC1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ST14 protein (Fragment).
GN Name=ST14;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny P.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Trinchieri J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Goughwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smaluk D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;

179 FAVDSIVAYSTDPQTTRIFRNSSCAVFLHSDGAVAKPSSPGFPDSDYPENARCLWTLRA 238
251 DADSVLSLTERSFDLASCDERGSDLVTVYNTLSNPEPHALVOLCGTYPYPSYNTLTFSSQN 310
239 NAGQIIHLHFKTFFMEKCKINGDGFVMVYDLSUPMEPRAQIRLCGIGYPPSYNTLTFPSSN 298
311 VLLITLITNTERRHFGFEATFFQLPRMSSCGRLRKAQGTFSPPYYPGHYPNIDCTWNI 370
299 VMLVTLVTDNVGKPPGFLAEFRQLPKSLCGSHIRDATGVITSPYFPAYPKTECIWDI 358
371 EVPNQHVKVSFKFPYLLPBGVPACTCPKDYVEINGEKYCGERSOFVTSNKNITVRPH 430
359 QVPGKKFVKLRFNMFLAEPGVPTKTDKFVEIEGQYCYGERKVFVWSNNSKISVRFV 418
431 SDQSYTDGFLAEVLSYSDSPCGQFCTRCGRCTRKELRCDCGMADCTDHSDELNCSDA 490
419 SDQSYADTGTADYLSYEPNPPDQFCTKRGCIIRLQCKDGMNDCEDFSEKCKCTFA 478
491 GHQFTC-KNPKCKPLFWYCDSDVNDGDNDSBQSCGSCPAQTFRCSNGKCLSKSQOQNGKDD 549
479 -QQFRCTDSKLCPSHFVCDGVNDGDNDSBELSCKPNSTFKCGNGKCFPDSQKCDRTDN 537
550 CGDGSDEASCPKVVNVVCTKHYRCLNGLCLSKGNPECDGKEDSDGSDKDD--CDCGLR 607
538 CGDGSDEADCGQVTTCTEYTKNNQCTTKNPECDGENDCTDGSDEASAKNCGR 597
608 SFTQARVVGCTDADEGEPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYS 667
598 PFTKSLIIGVADLGEFPQVSLHVKGSHTCGASLASPTLLISAACFOODHSMRYS 657
668 DPTQWTAFLGHDQSRSPAGVQBRRLKRIISHFFNFDDFTDYDIALLELEKPAEYSMV 727
658 DASLWATYGLGHDQAQNLNTKXNVVERKIRIWAHIGFNDNTYDNDIAVLELEKFPVDYDFI 717
728 RPICLPASHVFPAGKAIWVGWGHQYGGTGALILQKEIRVINTQTCENLLPOQITPR 787
718 QPICPESTDFPVPVKPIWVGWALKEGGAAVILQKAEIRVINQTECNLLDGLQTPR 777
788 MMCVFLSGGVDSQSGSGPLSSVEADGRIFQAGVSWGDCAGRNKPGVYVTRPLPRD 847
778 MLCAGFVSGGIDACQSGSGPLSSVDLNNKYLAVISWGEGGCARRNKPVTTRVMMRD 837
848 WIKENTGV 855
838 WIKDGTGL 845

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; BC018146; AALH1846.1; -;
 DR HSSP; Q9Y5Y6; IEAW.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.
 DR Pfam; PF00057; Ldl_recept_a; 4.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 1
 SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1858B319 CRC64;

Query Match 50.8%; Score 2379; DB 2; Length 422;
 Best Local Similarity 100.0%; Pred. No. 3.5e-155;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 434 SYTDGFLAEYLSYSSDPCPGQFTCTGTCRCIRKELRCGWDCTDHSDELNCSCDAGHQ 493
 DB 1 SYTDGFLAEYLSYSSDPCPGQFTCTGTCRCIRKELRCGWDCTDHSDELNCSCDAGHQ 60
 QY 494 FTCKNFKCKPLFWVCDVNDGNDSCGSCPAOTFCRCNGKCLSKSQCCNGKDDCGD 553
 DB 61 FTCKNFKCKPLFWVCDVNDGNDSCGSCPAOTFCRCNGKCLSKSQCCNGKDDCGD 120
 QY 554 SDEASCPKNVVTCTGTYRCLNGLCLSKGNPECDGKEDCSGDEKDCDGLRFTTQQA 613
 DB 121 SDEASCPKNVVTCTGTYRCLNGLCLSKGNPECDGKEDCSGDEKDCDGLRFTTQQA 180
 QY 614 RVVGGTDADEGEVQVQVSLHALGQGHICGASLISPNWLVSAACHYIDDRGFYSPTQWT 673
 DB 181 RVVGGTDADEGEVQVQVSLHALGQGHICGASLISPNWLVSAACHYIDDRGFYSPTQWT 240
 QY 674 AFLGLHDQSORANPGVQERLKRITISHPEFNDFTDYDIALLEKPAEYSWVRPICLP 733
 DB 241 AFLGLHDQSORANPGVQERLKRITISHPEFNDFTDYDIALLEKPAEYSWVRPICLP 300
 QY 734 DASHVFPAGKALVWTGWGHTQGGTGALILQGEIRVINTQTCENLLPQITPRMVCVGF 793
 DB 301 DASHVFPAGKALVWTGWGHTQGGTGALILQGEIRVINTQTCENLLPQITPRMVCVGF 360
 QY 794 LSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWMKENT 853
 DB 361 LSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWMKENT 420
 QY 854 GV 855
 DB 421 GV 422

RESULT 8

Q6DEV0 PRELIMINARY; PRT; 663 AA.
 AC Q6DEV0
 ID Q6DEV0
 DT 25-OCT-2004 (TREMBLrel. 28, Created)
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DE MGC89623 protein.
 GN Name=MGC89623;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC076994; AAH76994.1; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR009003; Pept_ser_Cys.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00057; Ldl_recept_a; 4.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN DOM; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 42.1%; Score 1969; DB 2; Length 663;
 Best Local Similarity 42.8%; Pred. No. 9.1e-127;
 Matches 363; Conservative 122; Mismatches 171; Indels 192; Gaps 6;
 QY 14 KDFGAGLKYNSRHEKVGLEEGVEFLPVNNVKKVKGPGRWVLAALVIGLLVLLGIG 73
 DB 2 KDSMMKYNRRPQSMNGFEGVEFLPAANTKRVKAGPKKGLAIFGVVIGALLSLTIG 61

```

QY 74 FLVWHLYQYRVRVYQKVFNGYMRITNENFVDAYENSTEFVSLASKVKDALKLLYSGVPP 133
DB 62 LLVWHFAYNAPVQKLYTGYLRANTQFVEAVENSTREFADLSVKVISTLTTLNGEKD 121
QY 134 LQYHKBESAVTAPSEGS----VIAYWSEFSIPQHLVVEAEARVMAEVRVMLPPRARSILKS 190
DB 122 IAPYLOQCSISAFSEGSNNVVGYYWSEFSVPAFEEAFAISE----LKLPTVNLQRQ 178
QY 191 FVTVSWAPPTSKVTQRTQDNCSFGLHARGVELMRTFTTPSPDPYPAHARCOWALRG 250
DB 179 FAVDSLVAITPQIARNFKNSCAFFLHSSAGVMTKFSPPGFTPTPPPNARCILWTLRA 238
QY 251 DADSVLSLTFRSGFDLASCDERGSIDLTVYNTLSPMEPHALVOLCTGYPSPSYNLTFHSSN 310
DB 239 DAGOMIRLAKFTKMEKKCANAGDFVMVYDLSPIEPRAQIRLCIGIYPPSYNLTFSSN 298
QY 311 VLLITLITNTERHGFETATFQLRPMSSCGRLKKAQGTFFNSPYPGHPNIDCTWNI 370
DB 299 VMLVTLVTDNVGFFGLAEFSQFPKTSLCGGYIRDASGVFTSPYFPFGHPKIECIWDI 358
QY 371 EVNNHVKVSKFFVLLPFGVAGTCTPKDYVEINGEKYCGERSOFVNTSNKITVRFH 430
DB 359 QVPDNKPKLRNFMFLAEPGVVPTKCTKDFEINGQYCGERKFFVNNSSNKSMSRFEV 418
QY 431 SDQSYDTGFLAELYSYSDSDPCPGQFTCRGRCIRKELCDGWADCTDHDDELNCSDA 490
DB 419 SDQSYDTGFLAELYSYSDSDPCPGQFTCRGRCIRKELCDGWADCTDHDDELNCSDA 478
QY 491 GHQFTCKN-KFCKPLFWVCDSDVNDGDSDEGSCCPAQTFRCNSGKCLSKSQOQNGKDD 549
DB 479 -LQFRCTNSKLCKPSYFVCDGVDGSDSDDELACQCPNNTYKCGNGKCIPEQKCDRTDN 537
QY 550 CGDGSDEASCPVWVTTCTHYRCINGLCLSKNGPCDCKGDCSDGSDSDE--KDCDCGLR 607
DB 538 CGDGSDEASCPVWVTTCTHYRCINGLCLSKNGPCDCKGDCSDGSDSDEISAKNCNGKR 597
QY 608 SFTQARVVGTHADEGEWPMQVSLHALQOQHICGASLSPNWLVSAAHCYIDDRGPRYS 667
DB 598 PTKSRIVGVNA----- 611
QY 668 DPTOWTAFGLHIDQSORAPGVQERRKRIISHPFNDFTFDYDIALLELKBPAEYSSMV 727
DB 612 ----- 611
QY 728 RPICLPDASHVFPAGKATVWTGHTQYGGTGALILQKEIRVINQITCENLLPQOITPR 787
DB 612 ----- 611
QY 788 MMCVGLSGGVDSQCGSDGGLSSVEADGRIFAGVWSGWDGCAORNKPGVYVTRLPFLRD 847
DB 612 -----DSGGPLSSVELNKNVYLAGIVSWGEGCARRNKPGVYVTRVAMWRD 655
QY 848 WIKENTGV 855
DB 656 WIRDKTGL 663

RESULT 9
ID Q6PF94 PRELIMINARY; PRT; 799 AA.
AC Q6PF94;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Tmpres6 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;

```

```

RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
EL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BC057674; AAH57674.1; -.
DR HSSP; P20160; 1AE5.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept Ser_Cys.
DR Pfam; PF00057; Ldl recept a; 1.
DR PRINTS; PR00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A6464D5288 CRC64;

```

```

Query Match 24.0%; Score 1124.5; DB 2; Length 799;
Best Local Similarity 31.6%; Pred. No. 1.2e-68;
Matches 266; Conservative 142; Mismatches 345; Indels 90; Gaps 25;
QY 33 EEGVE----FLPVNNVKVEKHGPGWVLAALVLLGLLGLGFLVWH-LOYR-DVRV 86
DB 20 EEAAEPGEKFKPKNTKR-----KNRDYRFTPLLVLALVAGVMLTFLGKAEVTV 74
QY 87 QKVFNGYMRITNENFVDAYENSTEFVSLASKVKDALKLLYSGVPPFLGYPHKEAVTAF 146
DB 75 SQVYSGSLRVNLNRHFSQDLGRRESIAFRSBSAKAQKMLQELVASTR-LGVYVNSSSVVSF 133
QY 147 SEGSVIAYWSEFSIPQHLVVEAEARVMAEVRV--MLPPRARSILKSPVTVSVVAFPTDSK 204
DB 134 GEGPLTCFFWFILDIPEY-----QRLTSPFVRELLVDELLSNSSTLASVKTEYEYDPE 188
QY 205 --TVQRTQDNCSFGLHARGVELMRTFTTPG--FP-DSPPFAHARCOWALRGDADSVLSLT 259
DB 189 GLVILEASVNDIVVLNSTGLCYRYSYVNPQVPLKGPDOQTTSCLHQLQPEDLMIKVR 248
QY 260 FRSFDLASCDERGSIDLTVYNTLSPMEPHALVOLCTGYPSPSYNLTFHSSQNVLITLN 319
DB 249 L-EWTRVDCRDR----VAMYDAAGPLEKRLITSVYGCGRQEPVMEVLASGSVMAVWVKKG 303

```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; AY240929; AAP69827.1; -	Db	146	GEGPLTCFFWFILDIPEY-----QRLTSPVVRRELLVDELLSNSSTLASKYTEVEVDP	200
CC	EMBL; AK004939; BAB23684.2; -	Qy	205	--TVQRTQDNCSFGLHARGVELMFTTPG--FP--DSPVPAHARQWALRGDADSVLSIT	259
CC	EMBL; BC029645; AAH29645.2; -	Db	201	GLVILEASNDIVLNLSTLGCYRYSYVNPQVLPKGPDDQQTSCIMHLQGPEDLMIKVR	260
CC	HSSP; P00760; LEZX.	Qy	260	FRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPSYNLTFHSSQNVLITLTN	319
DR	MEROPS; S01.308; -	Db	261	L-EMTRVDCDRD---VAMYDAAGPLEKRLITSVYGCQRQBPVMEVLASGVMVWVKKG	315
DR	MGD; MGI.1919003; TmpRSS6	Qy	320	TERRHPG-----EATFFQLPRMS--SCGORLKAQGTFSPPYGPYHNPINDCTWNIENVN	374
DR	GO; GO:0016021; C.integral to membrane; ISS.	Db	316	MHSYDPELLSVKSAVFQDCQVNLTEGLR-DTQGFLETPYPSYSPSTHCSMHLTVPS	374
DR	GO; GO:0004252; P.serine-type endopeptidase activity; ISS.	Qy	375	NOHKVSPKPYLLLEPGVPAGTCKPKDYVEINKEKYCGER-----SQFVVTSSNKKITVRF	429
DR	GO; GO:0001525; P.angiogenesis; ISS.	Db	375	LDYGLALWFDAYALRRQKNRLCTQGWMIQNRRLCGFRTLQPYAERIPWASDGVTFNF	434
DR	GO; GO:0042730; P.fibrinolysis; ISS.	Qy	430	HSQSYTDTGFLABYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDHSDELNCSD	489
DR	GO; GO:0007242; P.intracellular signaling cascade; ISS.	Db	435	TSQISLTGPGVQVYSLYNQSDPCGPEPLCSV-----	466
DR	GO; GO:0006508; P.proteolysis and peptidolysis; ISS.	Qy	490	AGHOPTCKNKECKPLFWVDSVNDGNSDQGCSCPAQTFRG--SNGKCLSKSQOCCNGKD	548
DR	InterPro; IPR000859; CUB.	Db	467	-----NGLCVF---ACDGIKDCPNGLDERNCVCRAM--EQCEDSTCISLPRVCDRQP	514
DR	InterPro; IPR002172; LDL receptor A.	Qy	549	DCGDSDEASCPKVVNVVCTKHTYRCLNGLCLSGNPECDGKEDCSDGSDKDCDGLRS	608
DR	InterPro; IPR009003; Pept Ser Cys	Db	515	DCLNGSDEEQOE--GVPCGTFTFQCDRSCVKKNPNECDGSDGSDGSDGSDGSDG	572
DR	InterPro; IPR001254; Peptidase_S1.	Qy	609	FTQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSD	668
DR	Pfam; PF00057; Ldl_recept_a; 2.	Db	573	LS--SRIVGGTVSSEGEPMQASLIQIRG--HICGALLIADRWVITAAHCFQED---SMAS	626
DR	Pfam; PF00089; Trypsin; 1.	Qy	669	PTQWTFGLHLDQQRGAPQVQERRLKRIISHPFNFDTFYDYDALLLEKPAEYSSMWR	728
DR	PRINTS; PR00722; CHYMOTRYPSIN.	Db	627	PKLWTFVLGKMRQNSR--WPGEVSPKVSRLFLPHYHEEDSHDYDVALQLDPVYVSATVR	685
DR	PROSITE; PS01180; CUB; 1.	Qy	729	PICLPDASHVFPACKATVVTGNGHTQYGGTTCALILQKEIRVINOCTTCENLLPOQITPRM	788
DR	PROSITE; PS01209; LDLR_1; 1.	Db	686	PVCLPASHFFEPQGHCHWITGWAQREGGVSNTLQKVQVQVLPQDLCEAYRVQVSPRM	745
DR	PROSITE; PS00668; LDLR_2; 3.	Qy	789	MCVGLSGGVDSQCGSGGLSSVEADGRIFQAGVSWGDCGCAQRNKGVTYTRLPFRDW	848
DR	PROSITE; PS0240; TRYPSIN_DOM; 1.	Db	746	LCAGYRKGKDACQDGGGLVLCREPSGRWFLAGLSVNGLCGRPNPFGVYTRVTRVINW	805
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.	Qy	849	IKE 851	
DR	PROSITE; PS00135; TRYPSIN_SER; 1.	Db	806	IQQ 808	
XW	Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor; Transmembrane.				
FT	DOMAIN 1 59				
FT	TRANSMEM 60 80				
FT	DOMAIN 81 811				
FT	DOMAIN 213 336				
FT	DOMAIN 323 440				
FT	DOMAIN 445 477				
FT	DOMAIN 478 514				
FT	DOMAIN 518 555				
FT	DOMAIN 565 799				
FT	ACT_SITE 617 617				
FT	ACT_SITE 668 668				
FT	ACT_SITE 762 762				
FT	CARBOHYD 138 138				
FT	CARBOHYD 184 184				
FT	CARBOHYD 216 216				
FT	CARBOHYD 338 338				
FT	CARBOHYD 433 433				
FT	CARBOHYD 453 453				
FT	CARBOHYD 518 518				
FT	CONFLICT 690 690				
SQ	SEQUENCE 811 AA; 90978 MW; 32EB3E7C3127801B CRC64;				
Query Match 24.0%; Score 1124.5; DB 1; Length 811;					
Best Local Similarity 31.6%; Pred. No. 1.3e-68;					
Matches 266; Conservative 142; Mismatches 345; Indels 90; Gaps 25;					
Qy	33	EEGVE-----FLPNVNVKVEKGPGRVWVLAVALGLLVLGIGFLVWVH-LQYR-DYRV	86		
Db	32	EEAAEPEGKFPKPKTKR-----KNRDYVRFTPLLLVLAALVSAGVWMLYFLGYKAEVTV	86		
Qy	87	OKVFNQYKRIITNPFVDYENSNSFVSLASKVDALKLYSGVFFLGPYHKSAVTAFA	146		
Db	87	SOVYSGSLRVLNHFQSQDLGRRESIAFRSESNAKQMLQELVASTR-LGTYNSSVVSF	145		
Qy	147	SEGSVIAYWBSFSPQHLVEAEVMAEERVV--MLPPRARSLKSFVVTVSVAFFPTDSK	204		

RESULT 11

ID	TMS6 HUMAN	STANDARD;	PRT;	811 AA.
AC	Q8IU80; Q8IU82; Q8IXV8;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Transmembrane protease, serine 6 (EC 3.4.21.-) (Matriptase-2).			
GN	Name=TMPRSS6;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RA	Hooper J.D., Quigley J.P.;			
RT	"TMPRSS6, a new type II transmembrane serine protease.";			
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;			
RA	Dunham I., Hunt A.R., Collins J.B., Bruskiewich R., Beare D.M.,			

RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Contoy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA McClay J., McLaren S., McMurray A.V., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuoyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Do A., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradehaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Letralle P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Haller L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budson M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shiruya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Franssion I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RA "The DNA sequence of human chromosome 22.";
RA Nature 402:489-495(1999).
RA [3]
RA SEQUENCE FROM N.A. (ISOFORM 2).
RA TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan E., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny P.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schlein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RA [4]
RA SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RX TISSUE=Fetal liver;
RX MEDLINE=22241917; PubMed=12149247; DOI=10.1074/jbc.M203007200;
RA Velasco G., Cal S., Quesada V., Sanchez L.M., Lopez-Ortin C.;
RA "Matricase-2, a membrane-bound mosaic serine proteinase predominantly
RA expressed in human liver and showing degrading activity against

RT extracellular matrix proteins.";
RL J. Biol. Chem. 277:37637-37646(2002).
RL [5]
RP REVIEW.
RX MEDLINE=22668120; PubMed=12784999; DOI=10.1023/A:1023003616848;
RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,
RA Bugge T.H., Antalis T.M.;
RA "Membrane anchored serine proteases: a rapidly expanding group of cell
RA surface proteolytic enzymes with potential roles in cancer";
RL Cancer Metastasis Rev. 22:237-258(2003).
CC -!- FUNCTION: May play a specialized role in matrix remodeling
CC processes in liver.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8IU80-1; Sequence=Displayed;
CC Swae=2;
CC IsoId=Q8IU80-2; Sequence=VSP_008379, VSP_008380;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Liver specific.
CC -!- SIMILARITY: Belongs to the peptidase S1 family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ319876; CAC85953.1; ALT_INIT.
CC EMBL; AY055383; AAL16413.1;
CC EMBL; AY055384; AAL16414.1;
CC EMBL; AL022314; -; NOT_ANNOTATED_CDS.
CC EMBL; BC039082; AAH39082.1;
CC HSSP; P00760; IEZX.
CC Genew; HGNC:16517; TMPSR56.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR02172; LDL_receptor_A.
CC InterPro; IPR009003; Pept_Ser_Cys.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00431; CUB; 1.
CC Pfam; PF00057; Ldl_recept_a; 2.
CC Pfam; PF00089; Trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00261; LDLRECEPTOR.
CC SMART; SM00042; CUB; 1.
CC SMART; SM00192; LDLa; 3.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS01180; CUB; 1.
CC PROSITE; PS01209; LDLRA_1; 1.
CC PROSITE; PS00068; LDLRA_2; 3.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Alternative splicing; Glycoprotein; Hydrolase; Repeat;
CC Serine protease; Signal-anchor; Transmembrane.
KW Cytoplasmic (Potential)
FT DOMAIN 56 76 Signal-anchor for type II membrane
FT TRANSMEM 56 76 protein (Potential).
FT FT Extracellular (Potential).
FT DOMAIN 77 811
FT DOMAIN 213 336 CUB 1.
FT DOMAIN 335 452 CUB 2.
FT DOMAIN 457 489 LDL-receptor class A 1.
FT DOMAIN 490 526 LDL-receptor class A 2.
FT DOMAIN 530 567 LDL-receptor class A 3.
FT DOMAIN 577 811 Serine protease.
FT ACT_SITE 617 617 Charge relay system (By similarity).
FT ACT_SITE 668 668 Charge relay system (By similarity).
FT

FT	ACT SITE	762	762	Charge relay system (by similarity).
FT	CARBOHYD	136	136	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	184	184	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	216	216	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	338	338	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	433	433	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	453	453	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	518	518	N-linked (GlcNAc. . .) (Potential).
FT	VARSPIC	409	461	LCGRILQYAEIRIYVAGTITINFTQSLTGVRVHY SLEAQVNLGRGASRGWCQACCP (in isoform 2).
FT	FTID-VSP	008379.		
FT	VARSPIC	462	811	Missing (in isoform 2).
FT	CONFLICT	116	116	/FTID-VSP 008380.
FT	SEQUENCE	811 AA;	89999 MW;	A -> V (in Ref. 4).
FT	SEQUENCE	811 AA;	89999 MW;	7BEF193F655DDE9D CRC64;
Qy	Query Match	24.0%;	Score 1124;	DB 1; Length 811;
Db	Best Local Similarity	31.0%;	Pred. No. 1.4e-68;	
Qy	Matches	276;	Conservative 138;	Mismatches 327; Indels 148; Gaps 28;
Qy	11	GGPKDFGAGLYNKRHEKVGLEGEVFLPVNNVKKVEKPGRMVLAALVGLLILVLL 70		
Db	20	GGQDGGDG-----EEAPEGFKACE-----DSKKARGYLRLVFLVLL--ALLVLA 66		
Qy	71	GIGFLVWH-LQYR-DVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKVKDKALKLY 128		
Db	67	SAGVLLWYFLGYKAEVWVSYSSGLVNLHFSDILTRRESSAFRSETAKQAKMLKELI 126		
Qy	129	SGVFLPGYKHSAVTFAPSEGSVIATYNSFSPHILVEAEAEVMAEVRVMLPPRARSL 188		
Db	127	TSTR-LGTYTNSSSVSYSGEGPLTCFFWFILQIPEH-----RRMLSPSEV--V 171		
Qy	189	KSFVTVSWAPTDSTKQVTDQNSCSFGLHARGVELMRP-----TTPGPDSPY- 238		
Db	172	QALLVEELSTVNSAAVYRAE-----YVDPEGLVILEASVKDIAALNSTGLCYRYSYV 227		
Qy	239	-----PAH--ARCOMALRGDADSVLSLTFPSFOLASCDERGSDDLVTVNTLSPME 286		
Db	228	GQGVLRKGPDLHASSCLMHLQPKDMLKRL-ENTLAECRDR-----LAMVDVAGPLE 282		
Qy	287	PHALVQL--CQTPPSYNLT-----FHSSQNVLLITLITNTERRHPGEATF 331		
Db	283	KRLITSVYGCRCQPVVEVLASGAIMAVWCKGLHSYDPPVLSV-----QPVV 331		
Qy	332	FQPRMSSCGRLKKAQGTNSPYPGHYPPNDICTWNIQVNNQHKVSKFKFYLLLEPG 391		
Db	332	FOACEVNLTLDNELDSQGLSTPTPYPSYSPQTHCSWHLTVPSLDYGLALWFDAYALRRQ 391		
Qy	392	VPAGTCPKDYVEINGEKVC-----ERSQFVVTNSNKITVRPHSDQSYDITDTGLAE 443		
Db	392	KYDLPCQGWNTQNRLLGLRILQYAEIRIPVATAG---ITINFTSQSLTGPGVRVH 448		
Qy	444	YLSYDSDPDCPGQTCRTGRGIRKELRCGWDGCTDHSDELNCSCDAGHOFCKNRFCKP 503		
Db	449	YGLYNQSDPCGPEGLCSV-----NGLCVP 472		
Qy	504	LFVWCVSDVNDGNSDQGSQCPAQTRFC--SNGKLSKSOQCNKGDCKGSGDEASCPKV 562		
Db	473	---ACDGVKDCPNGLDERNCVCRA--TPQCKEDSTCISLPKVCQDQPPCLNGSDEEQCE- 527		
Qy	563	NVVTCTKHTYCLNGLCLSGNPECDGKEDCDGSGDEKCDCLGRSFTROARVVGTDAD 622		
Db	528	-GVPCGTFITQEDRSCKVKNPQCDRQDGSDEEHCDGLQG---PSSRIIVGAVSS 584		
Qy	623	EGEWPQVSLHALGQGHICGASLISPNWLVSAAHYIDDRGFYRSDPTQWTAFLGLHDQS 682		
Db	585	EGEWPQVQASLQVRGR-HICGALLIADRWVITAAHCFQED---SWASTVLTWTLGKWNQ 640		
Qy	683	QRSAPGVQERRLKRIISHPFNFDTFDYDIALLEKPAEYSKVRPICLPDASHVFPAG 742		
Db	641	SR-WPGEVSPKVSRLLLHPYHEEDSHDYDVALQLDHPVVRSAARVPCVLPARSHPFEPG 699		

Qy	743	KAIWVTGHTQYGGTGALILQKEIRVINOTTENLLPOQITPRMCMCVGLSGVDSCQ 802
Db	700	LHCWITGAGLREGGSPISNALQKVDVQLIIPDLCEVYRVQVTRMLCAGYRKGGKDACQ 759
Qy	803	GDSGGPLSSVEADGRIPQAGVSWGDCGAQRNKGVVYTRLPFRDWIKE 851
Db	760	GDSGGPLVCKALSGRWFLAGLSWGLGCGRPNYFGVYTRITGVLSWISQ 808

RESULT 12

Q6UXD8	Q6UXD8	PRELIMINARY;	PRT;	802 AA.
AC	Q6UXD8;			
DT	05-JUL-2004	(T-EMBLrel. 27, Created)		
DT	05-JUL-2004	(T-EMBLrel. 27, Last sequence update)		
DT	05-JUL-2004	(T-EMBLrel. 27, Last annotation update)		
DE	PVAE354.			
GN	ORFNames=UNQ354;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22887296; PubMed=12975309; DOI=10.1101/gr.1293003;			
RA	Clark H.P., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,			
RA	Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,			
RA	Eaton D., Foster J.S., Grimaldi C., Gu Q., Hass P.E., Heldens S.,			
RA	Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,			
RA	Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,			
RA	Seshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,			
RA	Vandlen R., Watanabe C., Wieand D., Woods K., Xie M.H., Yansura D.,			
RA	Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,			
RA	Godowski P.,			
RT	"The secreted protein discovery initiative (SPDI), a large-scale			
RT	effort to identify novel human secreted and transmembrane proteins: a			
RT	bioinformatics assessment."			
RL	Genome Res. 13:2265-2270(2003).			
CC	-1- SIMILARITY: Belongs to peptidase family S1.			
DR	EMBL; AY358398; AAQ88764.1; -.			
DR	HSSP; P20160; IAE5.			
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.			
DR	GO; GO:0008233; F:peptidase activity; IEA.			
DR	GO; GO:0004295; F:trypsin activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR000859; CUB.			
DR	InterPro; IPR002172; LDL receptor A.			
DR	InterPro; IPR001254; Peptidase S1.			
DR	InterPro; IPR001314; Peptidase S1A.			
DR	InterPro; IPR009003; Pept_Ser_Cys.			
DR	Pfam; PF00431; CUB; 1.			
DR	Pfam; PF00057; Ldl_recept_a; 1.			
DR	Pfam; PF00089; Trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	PRINTS; PR00261; LDLRECEPTOR.			
DR	SMART; SM00192; LDLA; 3.			
DR	SMART; SM00020; Tryp_SPC; 1.			
DR	PROSITE; PS01180; CUB; 1.			
DR	PROSITE; PS01209; LDLRA_1; 1.			
DR	PROSITE; PS0068; LDLRA_2; 3.			
DR	PROSITE; PS0240; TRYPSIN DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN HIS; UNKNOWN_1.			
DR	PROSITE; PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease.			
SK	SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;			

Query Match	24.0%;	Score 1122;	DB 2; Length 802;
Best Local Similarity	31.0%;	Pred. No. 1.9e-68;	
Matches	276;	Conservative 137;	Mismatches 328; Indels 148; Gaps 28;
Qy	11	GGPKDFGAGLYNKRHEKVGLEGEVFLPVNNVKKVEKPGRMVLAALVGLLILVLL 70	
Db	20	GGQDGGDG-----EEAPEGFKACE-----DSKKARGYLRLVFLVLL--ALLVLA 66	

```
Db 11 GGGGDDGGD-----EAEPEGMFKACE-----DSKRKARGYLRLVPLFVLL--ALLVLA 57
Qy 71 GIGFLVWH-LOYR-DYRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKVKDAKLILY 128
Db 58 SAGVLLMYFLGKAEVVMVSGVYSGSLRVLNRHFSQDLTRRESSAFRSETAKAQMKELI 117
Qy 129 SGVPFLGPHKESAVTAFSEGSVIAYWGEFISIPQHLVEEAERVAERVMVLPARRSL 188
Db 118 TSTR-LGTYNSSSVYSGEGPLTCFFWFILOIPEH-----RRLMLSPEV--V 162
Qy 189 KSFWVTSVVAFTDSTKVTQTDNSCSFGLHARGVELMRP-----TTPGFPDSPY- 238
Db 163 QALLVEELLSTVNSSNAVPYRAE-----YEVDPGLVILEASVKDIAALNSTLGCYRYSYV 218
Qy 239 -----PAH--ARQWALRGDADSVLSLTFRSFDLASCDSRSDLVTVVNTLSPE 286
Db 219 GQGQVLRKGPDLHASSCLHMQPKDMLKRL-EWTLAECDR-----LAMYDVAGPLE 273
Qy 287 PHALVQL--CGTYPSPSYNLT-----FHSSQNVLLITLITWTERHFGFEATF 331
Db 274 KRLITSVYGCSPQEPVVEVLASGAIMAVVMKGLHSYDPPVLSV-----QPVV 322
Qy 332 FOLPRMSSCGRLRKAQGTNSPYYPGHYPNIDCTWNIEVNNQHVKSPPFFVLLBPG 391
Db 323 FOACEVNLTLNRLDSQGVLSLTPYFSPYSQTHCSWHLTVPSLDYGLAMFDAYALRRQ 382
Qy 392 VPAGTCPKDYVEINGEKYCG-----ERSQFVVTNSNKITVRHSDQSYTDTGFLAE 443
Db 383 KYDLPTQOGWTIQNRRLCGRLIOLPYAERI PVVATAG---ITINFTSQISLTGPGVRVH 439
Qy 444 YLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDHSDELNCSCDAGHQFTCKNFKCP 503
Db 440 YGLYNQSDPCPGFELCSV-----NGLCVP 463
Qy 504 LFWVCDVNDCCGNDQSCSPAOTERC-SNGKCLSKSQCCNGKDCGDSDEASCPKV 562
Db 464 ---ACGVKDCPNGLDERNVCRA-TFQCKEDSTCISLPKVCQDQDCLNGSDEBQOE- 518
Qy 563 NVVTCIKTYRCLNGCLSKNGPECDGKEDCDGSDGDEKDCDGLRSFTQARVVGTDAD 622
Db 519 -GVPCGTFQCEDRSCKVKNPQCDGRPDCHDGSDEHCDGLQG--PSSRIVGAVSS 575
Qy 623 EGEWPQVSLHALGQCHICGASLISPNMLVSAHCYIDDRGFRYSDPTQWTAFLGHDQS 682
Db 576 EGEWPQASLQVGR-HICGGALADRWWITAAHCFQED---SMASTVLTWTVFLGWQON 631
Qy 683 QRSAPGVQERLLRIISHPFNDFTDYDIALLEKPAEYSSMVRPCLPDASHVRPAG 742
Db 632 SR-WFGEVSVFVSRLLHPYHEEDSHDIDVALLQDHPVRSAAVRPCLPARSHFFBPG 690
Qy 743 KAIWTVGHGHTQYGGTGALILQKGBIRVINQTTCENTLLPQOITPRMCMVFLSGGVDSQC 802
Db 691 LHCWITWGALREGGPISNALQKVDVQLIPQDLCEAYRYQVTPRMLCAGYRKKGKDACQ 750
Qy 803 GDSGGLPSLEADGRIFOAGVSWGDCGCAQRNKPQVYTRLPFRDWIKE 851
Db 751 GDSGGLPVCKALSGRWFLAGLVSWGLGCGRPNYFGVYTRITGVISWIOQ 799
```

RESULT 13

Q6IC2

ID Q6IC2 PRELIMINARY; PRT; 824 AA.

AC Q6IC2;

DT 05-JUL-2004 (TreeBLrel. 27, Created)

DT 05-JUL-2004 (TreeBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TreeBLrel. 27, Last annotation update)

DE DJ1170K4.2 protein.

GN Name=DJ1170K4.2;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_taxid=9606;

RN [1]

```
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mallia M., Mokrab Y., Huckle E.J.,
RA Beare D.M., Dunham I.;
RL Submitted (MAY-2004) to the ENBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR HMBL; CR456446; CAG30332.1; -.
DR HSSP; P20160; 1AE5.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; Ldl_recept_a; 1.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 3.
DR PROSITE; PS01180; Tryp_SPC; 1.
DR PROSITE; PS01209; CUB; 1.
DR PROSITE; PS01209; LDLRA 1; 1.
DR PROSITE; PS00068; LDLRA 2; 3.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;
```

Query Match

Best Local Similarity 23.5%; Score 1102; DB 2; Length 824;

Matches 276; Conservative 137; Mismatches 328; Indels 170; Gaps 29;

```
Qy 11 GGPXDFGAGLKYNRHEKVGLEGEVFLPVNNVKKYKHPGRVWVLAALVILGLLLVLL 70
Db 11 GGGGDDGGD-----EAEPEGMFKACE-----DSKRKARGYLRLVPLFVLL--ALLVLA 57
Qy 71 GIGFLVWH-LOYR-DYRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKVKDAKLILY 128
Db 58 SAGVLLMYFLGKAEVVMVSGVYSGSLRVLNRHFSQDLTRRESSAFRSETAKAQMKELI 117
Qy 129 SGVPFLGPHKESAVTAFSEGSVIAYWGEFISIPQHLVEEAERVAERVMVLPARRSL 188
Db 118 TSTR-LGTYNSSSVYSGEGPLTCFFWFILOIPEH-----RRLMLSPEV--V 162
Qy 189 KSFWVTSVVAFTDSTKVTQTDNSCSFGLHARGVELMRP-----TTPGFPDSPY- 238
Db 163 QALLVEELLSTVNSSNAVPYRAE-----YEVDPGLVILEASVKDIAALNSTLGCYRYSYV 218
Qy 239 -----PAH--ARQWALRGDADSVLSLTFRSFDLASCDSRSDLVTVVNTLSPE 286
Db 219 GQGQVLRKGPDLHASSCLHMQPKDMLKRL-EWTLAECDR-----LAMYDVAGPLE 273
Qy 287 PHALVQL--CGTYPSPSYNLT-----FHSSQNVLLITLITWTERHFGFEATF 331
Db 274 KRLITSVYGCSPQEPVVEVLASGAIMAVVMKGLHSYDPPVLSV-----QPVV 322
Qy 332 FOLPRMSSCGRLRKAQGTNSPYYPGHYPNIDCTWNIEVNNQHVKSPPFFVLLBPG 391
Db 323 FOACEVNLTLNRLDSQGVLSLTPYFSPYSQTHCSWHLTVPSLDYGLAMFDAYALRRQ 382
Qy 392 VPAGTCPKDYVEINGEKYCG-----ERSQFVVTNSNKITVRHSDQSYTDTGFLAE 443
Db 383 KYDLPTQOGWTIQNRRLCGRLIOLPYAERI PVVATAG---ITINFTSQISLTGPGVRVH 439
Qy 444 YLSYDSSDPCPGQFTCTGRCIRKELRCGWDADCTDHSDELNCSCDAGHQFTCKNFKCP 503
Db 440 YGLYNQSDPCPGFELCSV-----NGLCVP 463
```


01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
Mus musculus adult male corpora quadrigenina cDNA, RIKEN full-length
enriched library, clone:B230219123 product:weakly similar to BLOOD
COAGULATION FACTOR XI
Name=Tmpres7;
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P.; Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P.; Shibata Y.; Hayatsu M.; Sugahara Y.; Shibata K.; Itoh M.;
RA Konno H.; Okazaki Y.; Muramatsu M.; Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K.; Itoh M.; Aizawa K.; Nagaoka S.; Sasaki N.; Carninci P.;
RA Konno H.; Akiyama J.; Nishi K.; Kitsuunai T.; Tashiro H.; Itoh M.;
RA Sumi N.; Ishii Y.; Nakamura S.; Hazama M.; Nishine T.; Harada A.;
RA Yamamoto B.; Matsumoto H.; Sakaguchi S.; Ikegami T.; Kashiwagi K.;
RA Fujitake S.; Inoue K.; Togawa Y.; Izawa M.; Ohara E.; Watahiki M.;
RA Yoneda Y.; Ishikawa T.; Ozawa K.; Tanaka T.; Matsuura S.; Kawai J.;
RA Okazaki Y.; Muramatsu M.; Inoue Y.; Kira A.; Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RA Adachi J.; Aizawa K.; Akimura T.; Arakawa T.; Bono H.; Carninci P.;
RA Fukuda S.; Furuno M.; Hanagaki T.; Hara A.; Hashizume W.
RA Hayashida K.; Hayatsu N.; Hiramoto K.; Hiraoka T.; Hirozane T.;
RA Hori F.; Imotani K.; Ishii Y.; Itoh M.; Kagawa I.; Kasukawa T.;
RA Katoh H.; Kawai J.; Kojima Y.; Kondo S.; Konno H.; Kouda M.; Koya S.;
RA Kurihara C.; Matsuyama T.; Miyazaki A.; Murata M.; Nakamura M.;
RA Nishi K.; Nomura K.; Numazaki R.; Ohno M.; Ohsato N.; Okazaki Y.;
RA Saito R.; Saichoh H.; Sakai C.; Sakai K.; Sakazume N.; Sano H.;
RA Sasaki D.; Shibata K.; Shinagawa A.; Shiraki T.; Sogabe Y.; Tagami M.;
RA Tagawa A.; Takahashi E.; Takaku-Akahira S.; Takeda Y.; Tanaka T.;
RA Tomaru A.; Toya T.; Yasunishi A.; Muramatsu M.; Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL: AK045663; BAC32448.1; --
DR HSP: P00760; 1EXZ.
DR MSKOPS; S01.072; --

MGD; MGI:2686594; Tmpres7.
GO; GO:0004263; P:chymotrypsin activity; IEA.
GO; GO:0008233; P:peptidase activity; IEA.
GO; GO:0004295; P:trypsin activity; IEA.
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
InterPro; IPR000859; CUB.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001314; Peptidase_S1A.
InterPro; IPR003003; Pept_Ser_Cys.
Pfam; PF00431; CUB; 1.
Pfam; PF00057; Ldl_recept_a; 3.
Pfam; PF00089; Trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
PRINTS; PR00261; LDLRECEPTOR.
SMART; SM00042; CUB; 1.
SMART; SM00192; LDLa; 3.
SMART; SM00202; Tryp_SPC; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01209; LDLRA_1; 1.
PROSITE; PS00068; LDLRA_2; 2.
PROSITE; PS02040; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 572 AA; 63757 MW; 21C0BC523B5F4301 CRC64;

Query Match 21.6%; Score 1011.5; DB 2; Length 572;
Best Local Similarity 36.5%; Pred. No. 4.8e-61;
Matches 226; Conservative 96; Mismatches 227; Indels 71; Gaps 21;

QY 244 COMALRGDADSVLSLTFERSFDLASCDBRGSDLVTVYNTLSPMEHALVOLCGTYPSPSYNL 303
DB 2 CHPKLVAVGVGLRLSIESIQL-EADNCITDSLTVDLSLLPIRSAILYRIC--EPTRTLM 58

QY 304 TFHSSQNVLLITLTNTERHPGFEATFFOLPMSSCGG-----RLRKAQTFSNPPYPG 358
DB 59 SFVSTNNMLVLKSPYVRRLAGIRAYFEVIPE-QKCESTILVKBIINSFEKGISSPYPS 117

QY 359 HYPNIDICTWNIENPVNQHVKVFKFVLEPGVPAGT'CPKDYVEINGEKYCGE--RSQF 416
DB 118 YYPEKCKTWTFTQ-T-SLSTIGIALKFPYNSITKSAKCEHGWEIHEHMYCGSMDET 176

QY 417 VVTSNSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCP-GQFTCTGRCITKELRCGWA 475
DB 177 IFRVPSPLVHIQCSRLSDKPLLVYGGYNTISQOCFAGSFRCSGLCVPAQRCDCGN 236

QY 476 DCTDHSDELNCSDAGHQFTCKNFKFPLFWVCDNVNDCGNSDEQSCGCPAQTFRCNSG 535
DB 237 DCFDESDELFCVT-----VXP-----ACNSSFR-QHG 263

QY 536 KCLSKSQCCNGKDDCGDGSDEASC'PKVNVVTCTKHYRCLNGLCLSKGNPCDCKEDCD 595
DB 264 PLV-----CDGFRDCEGDEQCNTR--SIPCTSRTEFKGNDICFRKNAQCGIVDCPD 316

QY 596 GSDEKDCDGLRSFTROARVVGTDADGEWPMQVLSLHAGQGHICGASLSPNWLVSAA 655
DB 317 GSDEEGCGCS-RSSSLFRLIRVGGSDSQEGTWPMQVLSLHFGVSAY-CGASVLSREWLSSAA 374

QY 656 HCVIDDRGFRYSPTQWTATFLGHDQSORS--APGVQERLKRITISHPPFNDFTFDYDIA 713
DB 375 HCF---HGNRLSDPTPTWTAHLGMVYVQNAKFI SP-----VRRIVVHEYNYSQTFDYDIA 425

QY 714 LLELE--KPAEYSMSVRPICLPDASHVFPAGKAIWVTGWGHT-QYGGTGALILQKGI 770
DB 426 LLQLSTAWPETLKLQIQICIPAGQKVRSEKCNWVGRRHEADSKGSPVLOQAEVEL 485

QY 771 INQTTCCNLAPQOITPRMCMVGLSGVSDSCQDGGPLS-SVEADGRIFQAGVVSNGDG 829
DB 486 IDQTVCVSTY-GIITSRLCAGVMSGKSDACKGDSGGPLSCRRKSDGKWLITGIVSGHG 544

QY 830 CAQRNKGVTYTRLPFDWI 849
DB 1 : | | | | | : | | | | |

Fri Sep 23 14:45:56 2005

us-09-421-213-2.rup

Page 17

Db 545 CGRNPFGVYTRYSSFVPWI 564

Search completed: September 23, 2005, 12:59:09
Job time : 184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:46:25 ; Search time 82 Seconds
(without alignments)
4032.684 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGSDRRKGGGPKDFGAGL.....PGVYTRFLFRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: geneseq1980s:*
- 2: geneseq1990s:*
- 3: geneseq2000s:*
- 4: geneseq2001s:*
- 5: geneseq2002s:*
- 6: geneseq2003as:*
- 7: geneseq2003bs:*
- 8: geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4681	100.0	855	2	AAY06671 Tumour an
2	4681	100.0	855	4	AAB98500 Human TAD
3	4681	100.0	855	4	AAB06930 Human mem
4	4681	100.0	855	5	AAO22929 Type II t
5	4681	100.0	855	5	ADII16816 Human NOV
6	4681	100.0	855	5	ADII16882 Human NOV
7	4681	100.0	855	5	ADII16875 Human NOV
8	4681	100.0	855	6	ABP56619 Human mem
9	4681	100.0	855	6	AAO30146 Human mem
10	4681	100.0	855	6	AAE29820 Human mem
11	4681	100.0	855	6	AAE29791 Human mem
12	4681	100.0	855	6	ABP72376 Transmem
13	4681	100.0	855	7	ADBP97551 Human MTS
14	4681	100.0	855	7	ADII10371 Human cel
15	4681	100.0	855	8	ADG65326 Human MTS
16	4681	100.0	855	8	ADI28861 Human mat
17	4681	100.0	855	8	ADJ46895 Human tra
18	4676	99.9	855	3	AAB19552 Human mat
19	4676	99.9	855	4	AAAB35465 Human mem
20	4676	99.9	855	5	ADII16817 Human NOV
21	4676	99.9	855	5	ADII16883 Human NOV
22	4676	99.9	855	5	ADII16876 Human NOV
23	4676	99.9	855	5	ADN39867 Cancer/an
24	4676	99.9	855	8	ADN04754 Antipsori
25	4676	99.9	855	8	ADP23334 PRO polyp

ALIGNMENTS

RESULT 1

RAY06671
ID RAY06671 standard; protein; 855 AA.

XX AC RAY06671;

DT 09-NOV-1999 (first entry)

DE Tumour antigen derived gene-15 (TAGD-15) protein.

XX Tumour antigen derived gene-15; TAGD-15; serine protease; human;
KW breast cancer; ovary cancer; carcinoma; diagnosis.

XX Homo sapiens.

Key	Location/Qualifiers
FT Domain	1..54 /note= "cytoplasmic domain"
FT Domain	55..213 /note= "transmembrane domain"
FT Modified-site	109..111 /note= "Asn is N-glycosylated"
FT Region	214..447 /note= "CUB repeat"
FT Modified-site	302..304 /note= "Asn is N-glycosylated"
FT Region	453..602 /note= "ligand-binding repeat (class A motif)"
FT Region	481..483 /note= "conserved SDE motif"
FT Region	518..520 /note= "conserved SDE motif"
FT Region	554..556 /note= "conserved SDE motif"
FT Region	597..599 /note= "conserved SDE motif"
FT Cleavage-site	614..615 /note= "catalytic domain"
FT Domain	615..855 /note= "catalytic domain"

XX WO9942120-A1.

XX 26-AUG-1999.

XX 18-FEB-1999; 99WO-US003436.

XX 20-FEB-1998; 98US-00027337.

XX

PA (UYAR-) UNIV ARKANSAS
XX O'brien TJ, Tanimoto H;
XX WPI; 1999-527418/44.
DR N-PSDB; AAX87815.
XX A new extracellular serine protease for diagnosis of neoplastic disease.
XX Claim 3; Fig 10; 7lpp; English.
XX The present sequence represents a novel human extracellular serine
CC protease, termed tumor antigen derived gene-15 protein (see AAX06671),
CC that is overexpressed in breast and ovarian carcinomas. The TAGD-15 gene
CC (see AAX87815) can be used as a diagnostic and therapeutic target in
CC ovarian carcinoma and other carcinomas including breast, prostate, lung
CC and colon. The ligand binding domains of TAGD-15 may be valuable in the
CC uptake of specific molecules into tumor cells. The invention also
CC provides: a vector that is capable of expressing DNA encoding TAGD-15
CC protein; host cells selected from bacterial cells (especially Escherichia
CC coli), mammalian cells, plant cells and insect cells; and a method of
CC detecting expression of TAGD-15 protein using a hybridisation probe
XX
XX Sequence 855 AA;
SQ
Query Match 100.0%; Score 4681; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.1e-307;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGRWVLA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGRWVLA 60
QY 61 VLIGLLVLLGIGFLVHLYQYDVRVQKVFNGVMRTNENFVDAYENSSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGFLVHLYQYDVRVQKVFNGVMRTNENFVDAYENSSTEFVSLASKV 120
QY 121 KDALKLLYSGVPLFGPYKHESAVTAFSEGSVIAYTWSEFSIPQHLVEEAERYMAEERVVM 180
Db 121 KDALKLLYSGVPLFGPYKHESAVTAFSEGSVIAYTWSEFSIPQHLVEEAERYMAEERVVM 180
QY 181 LPPRARSLSKSFVTSVVAFTDSKTVQRTQDSCSFLGHARGVELMFTTFCPPSPYPA 240
Db 181 LPPRARSLSKSFVTSVVAFTDSKTVQRTQDSCSFLGHARGVELMFTTFCPPSPYPA 240
QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGLVTVNTLSMPHEHALVOLCGTTPPS 300
Db 241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGLVTVNTLSMPHEHALVOLCGTTPPS 300
QY 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY 361 PNINDCTWNIENPNQHVSKFPFVLLPFGVPAGTCDPKDYVEINGEKYCGERSQFVTS 420
Db 361 PNINDCTWNIENPNQHVSKFPFVLLPFGVPAGTCDPKDYVEINGEKYCGERSQFVTS 420
QY 421 NSNKITVRPHSDQSYTDGTFLAEYLSYSSDDPCPGQTCRTGCRIRKELRCGDWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDGTFLAEYLSYSSDDPCPGQTCRTGCRIRKELRCGDWADCTDH 480
QY 481 SDLNCSCDAGHQFTCKNFKCPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDLNCSCDAGHQFTCKNFKCPLFWVCDSDVNDGDNDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SOQCNKGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
Db 541 SOQCNKGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
QY 601 CDCCLGRSFTROARVVGGTADGEPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 CDCCLGRSFTROARVVGGTADGEPWQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFYSDPTQWTAFLGLHDQSRAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEXP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSRAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEXP 720
QY 721 AYSMSVRPICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 AYSMSVRPICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
QY 781 PQQITPRMVCVGLSGGVDSCQSDGGPLSSVEADGRIFQAGVSVSWGDCGCAORNKPGV 840
Db 781 PQQITPRMVCVGLSGGVDSCQSDGGPLSSVEADGRIFQAGVSVSWGDCGCAORNKPGV 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 2
AAB98500
ID AAB98500 standard; protein; 855 AA.
XX AAB98500;
XX 03-AUG-2001 (first entry)
XX Human TAGD-15.
XX Human; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer;
KW tumour antigen-derived gene 15; extracellular serine protease.
XX Homo sapiens.
XX WO200129056-A1.
XX 26-APR-2001.
XX 20-OCT-2000; 2000MO-US029095.
XX 20-OCT-1999; 99US-00421213.
XX (UYAR-) UNIV ARKANSAS.
XX O'brien TJ, Tanimoto H;
XX WPI; 2001-381031/40.
XX N-PSDB; AAX23601.
XX Novel extracellular serine protease, termed tumor antigen-derived gene 15
PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,
PT treatment, prevention of cancer, particularly breast, ovarian cancer.
XX Claim 11; Fig 2; 130pp; English.
XX The present sequence represents human tumour antigen-derived gene 15
CC (TAGD-15) protein. TAGD-15 is an extracellular serine protease. It was
CC found that TAGD-15 is over-expressed in ovarian tumours. TAGD-15 protein
CC or its fragments of 9-20 residues that lack TAGD-15 protease activity are
CC useful for vaccinating an individual against TAGD-15, having, suspected
CC of having or at risk of getting cancer. Furthermore, the TAGD-15 gene can
CC be used as a diagnostic or therapeutic target in cancer
XX
SQ Sequence 855 AA;
Query Match 100.0%; Score 4681; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.1e-307;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGRWVLA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKKVKGHGRWVLA 60
QY 61 VLIGLLVLLGIGFLVHLYQYDVRVQKVFNGVMRTNENFVDAYENSSTEFVSLASKV 120

Db	61	VLIGLLLVLLGIGFLVWHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV	120
Qy	121	KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYTWSBFSIPQHLVEAEARVMAEERVVM	180
Db	121	KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYTWSBFSIPQHLVEAEARVMAEERVVM	180
Qy	181	LPPEARSLKSFVVTSVVAPPTDSKTQRTQDNCSFGLHARGVELMRFTTGGFPDPSYP	240
Db	181	LPPEARSLKSFVVTSVVAPPTDSKTQRTQDNCSFGLHARGVELMRFTTGGFPDPSYP	240
Qy	241	HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVYNTLSMPHEHALVOLCGTTPPS	300
Db	241	HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVYNTLSMPHEHALVOLCGTTPPS	300
Qy	301	YNLTFFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY	360
Db	301	YNLTFFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY	360
Qy	361	PPNIDCTWNIETVPNNQHVSKFPIYLLPBPVAGTCTPKDYVEINGEKYCGERSQFVVT	420
Db	361	PPNIDCTWNIETVPNNQHVSKFPIYLLPBPVAGTCTPKDYVEINGEKYCGERSQFVVT	420
Qy	421	NSNKITVRPHSDQSYTDGFLAEYLSDSSDPCGQFTCRGRCIRKELRCOGWADCTDH	480
Db	421	NSNKITVRPHSDQSYTDGFLAEYLSDSSDPCGQFTCRGRCIRKELRCOGWADCTDH	480
Qy	481	SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGNSDEQCSCPAQTFRCNSGKCLSK	540
Db	481	SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGNSDEQCSCPAQTFRCNSGKCLSK	540
Qy	541	SOQCNGKDDCGDSDEASCPKVVNTCTKHTYRCLNGLCLSKNPECDGKEDCDGSDSEK	600
Db	541	SOQCNGKDDCGDSDEASCPKVVNTCTKHTYRCLNGLCLSKNPECDGKEDCDGSDSEK	600
Qy	601	DCDCGLRFTQARVVGTDGEGWPQVSIHALGQGHICGASLISPNWLVSAAHCYID	660
Db	601	DCDCGLRFTQARVVGTDGEGWPQVSIHALGQGHICGASLISPNWLVSAAHCYID	660
Qy	661	DRGRYSPTQWTAFLGLHDQSAPGVQERRLKRIISHPFNDFTPDYDIALLEK	720
Db	661	DRGRYSPTQWTAFLGLHDQSAPGVQERRLKRIISHPFNDFTPDYDIALLEK	720
Qy	721	AEYSMSVRPICLEDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL	780
Db	721	AEYSMSVRPICLEDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL	780
Qy	781	POQITPRMCMVGHLSGGVDSQGDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP	840
Db	781	POQITPRMCMVGHLSGGVDSQGDGGLSSVEADGRIFQAGVSWGDCGCAQRNKP	840
Qy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855
RESULT 3			
ID	AAE06930		
XX	AAE06930 standard; protein; 855 AA.		
AC	AAE06930;		
XX			
DT	16-Oct-2001 (first entry)		
XX			
DE	Human membrane-type serine protease (MTSP) 1.		
XX			
KW	Human; transmembrane serine protease; membrane-type serine protease;		
KW	MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;		
KW	lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;		
XX	matriptase.		
OS	Homo sapiens.		
XX			
FH	Key Location/Qualifiers		

FT	Domain	615..855	
FT	XX	/label= Protease_domain	
PN	XX	WO200157194-A2.	
XX	XX	09-AUG-2001.	
PD	XX		
XX	XX	02-FEB-2001; 2001WO-US003471.	
XX	XX	03-FEB-2000; 2000US-0179982P.	
PR	XX	18-FEB-2000; 2000US-0193542P.	
PR	XX	22-JUN-2000; 2000US-0213124P.	
PR	XX	26-JUL-2000; 2000US-0220970P.	
PR	XX	08-SEP-2000; 2000US-00657986.	
PR	XX	22-SEP-2000; 2000US-0234840P.	
PA	XX	(CORV-) CORVAS INT INC.	
XX	XX		
PI	XX	Madison EL, Ong EO, Yeh J;	
XX	XX	WPI; 2001-488877/53.	
DR	XX	N-PSDB; AAD13113.	
XX	XX		
FT	XX	Novel single chain polypeptide comprising protease domain of type-II	
FT	XX	membrane-type serine protease or its catalytically active portion useful	
FT	XX	for treating and preventing cancer and tumor.	
PT	XX	Claim 12; Page 195-197; 256pp; English.	
XX	XX		
PS	XX	The invention relates to transmembrane serine proteases and their	
XX	XX	corresponding nucleotides and the protease domain of a type-II membrane-	
CC	XX	type serine protease (MTSP). MTSP is useful for identifying compounds	
CC	XX	that modulate or inhibit its proteolytic activity and for formulating a	
CC	XX	medicament for treating neoplastic disease. MTSP and its corresponding	
CC	XX	nucleotides are useful in preventing or treating tumours or cancers such	
CC	XX	as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in	
CC	XX	diagnostics and in hybridisation assays. MTSP is useful as a diagnostic	
CC	XX	marker for tumour development, growth and/or progression and as	
CC	XX	immunogens to generate antibodies that specifically bind to it. MTSP DNA	
CC	XX	is useful in a yeast two-hybrid system and in gene therapy. The present	
CC	XX	sequence is human MTSP1 protein (also called matriptase)	
XX	XX		
SQ	XX	Sequence 855 AA;	
Query Match 100.0%; Score 4681; DB 4; Length 855;			
Best Local Similarity 100.0%; Pred. No. 2.1e-307;			
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MGSDRARKGGGPKDFGAGLKYNRSRHEKVNGLBEGVEFLPVNNVKYKXGPGRWVLLAA	60
Db	1	MGSDRARKGGGPKDFGAGLKYNRSRHEKVNGLBEGVEFLPVNNVKYKXGPGRWVLLAA	60
Qy	61	VLIGLLLVLLGIGFLVWHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV	120
Db	61	VLIGLLLVLLGIGFLVWHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV	120
Qy	121	KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYTWSBFSIPQHLVEAEARVMAEERVVM	180
Db	121	KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYTWSBFSIPQHLVEAEARVMAEERVVM	180
Qy	181	LPPEARSLKSFVVTSVVAPPTDSKTQRTQDNCSFGLHARGVELMRFTTGGFPDPSYP	240
Db	181	LPPEARSLKSFVVTSVVAPPTDSKTQRTQDNCSFGLHARGVELMRFTTGGFPDPSYP	240
Qy	241	HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVYNTLSMPHEHALVOLCGTTPPS	300
Db	241	HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVYNTLSMPHEHALVOLCGTTPPS	300
Qy	301	YNLTFFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY	360
Db	301	YNLTFFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY	360
Qy	361	PPNIDCTWNIETVPNNQHVSKFPIYLLPBPVAGTCTPKDYVEINGEKYCGERSQFVVT	420

Db 361 PPNIDCTWNIIEVFNHVKVSKFVYLLPFGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
 Qy 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Db 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Qy 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Qy 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Db 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Qy 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660
 Db 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660
 Qy 661 DRGFRYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Qy 721 AYSYSSMVRPCLIPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE 780
 Db 721 AYSYSSMVRPCLIPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE 780
 Qy 781 PQOITPRMVCVGLSGGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYT 840
 Db 781 PQOITPRMVCVGLSGGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYT 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 4
 ID AAO22929 standard; protein; 855 AA.
 AC AAO22929;
 XX AAO22929;
 DT 12-DEC-2002 (first entry)
 DE Type II transmembrane serine protease 1 protein SEQ ID No 2.
 KW Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
 KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
 KW malignant; enzyme.
 XX Homo sapiens.
 XX WO200272786-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US007903.
 XX 13-MAR-2001; 2001US-0275592P.
 XX (CORV-) CORVAS INT INC.
 XX Madison EL, Ong EO;
 XX WPI; 2002-732827/79.
 XX N-PSDB; AAL53444.
 XX New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
 PT neoplastic diseases, monitoring tumor progress or therapeutic
 PT effectiveness, or identifying MTSP7 modulators for treating tumors or
 PT cancers.
 XX Disclosure; Page 172-174; 184pp; English.

XX The invention relates to a purified single or two-chain polypeptide,
 CC which comprises the protease domain of a type-II membrane-type serine
 CC protease 7 (MTSP7) or its catalytically active portion. The polypeptide
 CC comprising MTSP7 is useful for detecting or diagnosing a neoplastic
 CC disease, a pre-malignant lesion, a malignancy or other pathological
 CC condition in a subject. This polypeptide is also useful for monitoring
 CC tumour (e.g. tumor of the breast, cervix, prostate, lung, ovary or
 CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
 CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
 CC for treating or preventing a neoplastic disease, or tumour initiation,
 CC growth or progression, or a (pre-)malignant condition. The polypeptide or
 CC polynucleotide is also useful for identifying modulators of MTSP7, which
 CC may be used to treat cancers or tumours. This sequence represents a
 CC protein of the type-II membrane-type serine protease 1 relating to the
 CC invention
 XX
 SQ Sequence 855 AA;
 Query Match 100.0%; Score 4681; DB 5; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNKLESGVEFLPVNNYKVKVKGHRWVLA 60
 Db 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNKLESGVEFLPVNNYKVKVKGHRWVLA 60
 Qy 61 VLIGLLVLLGIGFLVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 Db 61 VLIGLLVLLGIGFLVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 Qy 121 KDALKLLYSVPFLGYPYHKS AVTAFSEGSVIAYWSEFSIPQHLVEFAERVMAEERVVM 180
 Db 121 KDALKLLYSVPFLGYPYHKS AVTAFSEGSVIAYWSEFSIPQHLVEFAERVMAEERVVM 180
 Qy 181 LPPRARSLSKSVTVSVAFPTDSKTQVORTDONS CSFGLHARGVLMRTTTPGFPDSPA 240
 Db 181 LPPRARSLSKSVTVSVAFPTDSKTQVORTDONS CSFGLHARGVLMRTTTPGFPDSPA 240
 Qy 241 HARCQWALRGDADSVLSITFRSFDLASCDSRGSDLVTVTNTLSPMEPHALVQLCCTYPPS 300
 Db 241 HARCQWALRGDADSVLSITFRSFDLASCDSRGSDLVTVTNTLSPMEPHALVQLCCTYPPS 300
 Qy 301 YNLTFHSSQNVLLITLTNTERRHPGPEATFPQLPRMSSCGGRLRKAQGTNSPYPGHY 360
 Db 301 YNLTFHSSQNVLLITLTNTERRHPGPEATFPQLPRMSSCGGRLRKAQGTNSPYPGHY 360
 Qy 361 PPNIDCTWNIIEVFNHVKVSKFVYLLPFGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
 Db 361 PPNIDCTWNIIEVFNHVKVSKFVYLLPFGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
 Qy 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Db 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Qy 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Qy 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Db 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Qy 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660
 Db 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660
 Qy 661 DRGFRYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Qy 721 AYSYSSMVRPCLIPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE 780
 Db 721 AYSYSSMVRPCLIPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE 780

QY 121 KDALKLLYSVGFPLGPHKESAVTAFSEGSVIAYWSEFSIQHLVEAEARVMAERVVM 180
 DB 121 KDALKLLYSVGFPLGPHKESAVTAFSEGSVIAYWSEFSIQHLVEAEARVMAERVVM 180
 QY 181 LPPRARSLKS FVVTSSVVAFTDSTKVQRTQDNCSFGLHARGVELMRFTTPGFPDPSYPA 240
 DB 181 LPPRARSLKS FVVTSSVVAFTDSTKVQRTQDNCSFGLHARGVELMRFTTPGFPDPSYPA 240
 QY 241 HARCQWALRGDADSVLSLFRSFDLASCDERSGLVTVNTLSMPHEHALVOLCGTYPPS 300
 DB 241 HARCQWALRGDADSVLSLFRSFDLASCDERSGLVTVNTLSMPHEHALVOLCGTYPPS 300
 QY 301 YNLTFHSSQVLLITLITNTRRHGCFEATFFQLPRMSSCGGLRKAQGTFSPIYPGHY 360
 DB 301 YNLTFHSSQVLLITLITNTRRHGCFEATFFQLPRMSSCGGLRKAQGTFSPIYPGHY 360
 QY 361 PNIDCTWNIENPNQHVKSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGRSQFVVT 420
 DB 361 PNIDCTWNIENPNQHVKSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGRSQFVVT 420
 QY 421 NSNKLTVRPHSDQSYTDTGFLAEYLSYDSSDPCQFTCTGRCIRKELRCGWDADCTDH 480
 DB 421 NSNKLTVRPHSDQSYTDTGFLAEYLSYDSSDPCQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSCDAGHFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSCPAQTFRCNSGKCLSK 540
 DB 481 SDELNCSCDAGHFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSCPAQTFRCNSGKCLSK 540
 QY 541 SQQCKGKDDCGSDGSEASCPKVVVYCTKHYTRCLNGLCLSKNPECDGKEDCSGSD 600
 DB 541 SQQCKGKDDCGSDGSEASCPKVVVYCTKHYTRCLNGLCLSKNPECDGKEDCSGSD 600
 QY 601 DCDCGLRFTROARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 DB 601 DCDCGLRFTROARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFRYSPTQWTAFLGLHDQSQSAPGVQERLKRILSHPPFNFTFDYDIALLELEKP 720
 DB 661 DRGFRYSPTQWTAFLGLHDQSQSAPGVQERLKRILSHPPFNFTFDYDIALLELEKP 720
 QY 721 ABEYSWVRPILCPDASHVFPAGKATWVTGWGHTQYGGTGALILQGEIRVINQTCENLL 780
 DB 721 ABEYSWVRPILCPDASHVFPAGKATWVTGWGHTQYGGTGALILQGEIRVINQTCENLL 780
 QY 781 PQOITPRMCMVGFSLGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCQORNKPGVYT 840
 DB 781 PQOITPRMCMVGFSLGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCQORNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855

RESULT 6

AD116882
 ID AD116882 standard; protein; 855 AA.

XX AD116882;

XX 15-APR-2004 (first entry)

XX Human NOVX protein homologue SeqID 418.

XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
 KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
 KW Alzheimer's disease; infection; str.

OS Homo sapiens.

XX WO200268649-A2.

XX 06-SEP-2002.
 XX 31-JAN-2002; 2002WO-US002785.
 XX 31-JAN-2001; 2001US-0265395P.
 PR 31-JAN-2001; 2001US-0265412P.
 PR 31-JAN-2001; 2001US-0265514P.
 PR 31-JAN-2001; 2001US-0265517P.
 PR 02-FEB-2001; 2001US-0266406P.
 PR 05-FEB-2001; 2001US-0266767P.
 PR 07-FEB-2001; 2001US-0266975P.
 PR 07-FEB-2001; 2001US-0267057P.
 PR 08-FEB-2001; 2001US-0267459P.
 PR 09-FEB-2001; 2001US-0267823P.
 PR 15-FEB-2001; 2001US-0268974P.
 PR 26-FEB-2001; 2001US-0271664P.
 PR 27-FEB-2001; 2001US-0271839P.
 PR 27-FEB-2001; 2001US-0271855P.
 PR 02-MAR-2001; 2001US-0272788P.
 PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275989P.
 PR 15-MAR-2001; 2001US-0276448P.
 PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276397P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 26-MAR-2001; 2001US-0278778P.
 PR 29-MAR-2001; 2001US-0279882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282922P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0288327P.
 PR 03-MAY-2001; 2001US-0288504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0298959P.
 PR 19-JUN-2001; 2001US-0299324P.
 PR 13-AUG-2001; 2001US-0312020P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.
 PR 31-AUG-2001; 2001US-0316447P.
 PR 07-SEP-2001; 2001US-0318115P.
 PR 07-SEP-2001; 2001US-0318118P.
 PR 12-SEP-2001; 2001US-0318740P.
 PR 19-SEP-2001; 2001US-0323379P.
 PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.

Tchernev VT, Spytek KA, Zerhusen BD, Patturajan M, Shinkets RA;
 Li L, Gangolli EA, Padigar M, Anderson DW, Rastelli L, Miller CE;
 Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;
 Furtak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;

WPI; 2002-706998/76.

New NOVX polypeptides and nucleic acids, useful for preventing or
 treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 pharmacogenomics.

26-MAR-2001; 2001US-0278775P.
 26-MAR-2001; 2001US-0278778P.
 29-MAR-2001; 2001US-0279882P.
 29-MAR-2001; 2001US-0279884P.
 30-MAR-2001; 2001US-0280147P.
 11-APR-2001; 2001US-0282992P.
 11-APR-2001; 2001US-0283083P.
 20-APR-2001; 2001US-0285133P.
 23-APR-2001; 2001US-0285749P.
 03-MAY-2001; 2001US-0288327P.
 03-MAY-2001; 2001US-0288504P.
 29-MAY-2001; 2001US-0294047P.
 30-MAY-2001; 2001US-0294473P.
 08-JUN-2001; 2001US-0296964P.
 18-JUN-2001; 2001US-0298959P.
 19-JUN-2001; 2001US-0299324P.
 13-AUG-2001; 2001US-0312020P.
 16-AUG-2001; 2001US-0312889P.
 16-AUG-2001; 2001US-0312908P.
 21-AUG-2001; 2001US-0313390P.
 28-AUG-2001; 2001US-0315470P.
 31-AUG-2001; 2001US-0316447P.
 07-SEP-2001; 2001US-0318115P.
 07-SEP-2001; 2001US-0318118P.
 12-SEP-2001; 2001US-0318740P.
 19-SEP-2001; 2001US-0323379P.
 18-OCT-2001; 2001US-0330245P.
 18-OCT-2001; 2001US-0330308P.
 14-NOV-2001; 2001US-0332701P.
 (CURA-) CURAGEN CORP.
 Tchernev VT, Spytek KA, Zerhusen BD, Patturajan M, Shimkets RA, Li L, Gangolli EA, Padigaru M, Anderson DW, Raestelli L, Miller CE, Gerlach VL, Taupier RJ, Gusev VV, Colman SD, Wolenc AR, Pena CE, Furtak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE, WPI; 2002-706998/76.
 New NOVX polypeptides and nucleic acids, useful for preventing or treating NOVX-associated disorders, e.g. cancer, cardiomyopathy, atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or pharmacogenomics.
 Disclosure; SEQ ID NO 411; 1498pp; English.
 This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism.
 Specifically, it refers to the use of biologically active fragments for diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, antiarthritic, hepatotropic, neuroprotective, nootropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example, neurogenesis, wound healing and angiogenesis. The nucleic acids are also used as in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. This polypeptide is a homologue of a human NOVX protein of the invention.

SQ Sequence 855 AA;			
Query Match 100.0%; Score 4681; DB 5; Length 855;			
Best Local Similarity 100.0%; Pred. No. 2.1e-307;			
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MGSDRARKGGGPKD	FGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKEHGGWVLLAA 60
Db	1	MGSDRARKGGGPKD	FGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKEHGGWVLLAA 60
Qy	61	VLIIGLLVLLIGL	GVWHLQYRDYRVQKVFNGYWRITNENFVDAYENSNSTEFVSLASKV 120
Db	61	VLIIGLLVLLIGL	GVWHLQYRDYRVQKVFNGYWRITNENFVDAYENSNSTEFVSLASKV 120
Qy	121	KDALKLYSGVPLG	YPYHKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEERVMAEERVVM 180
Db	121	KDALKLYSGVPLG	YPYHKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEERVMAEERVVM 180
Qy	181	LPPRARSLSKGFV	TSVVAFPDTSKTQVQTDNSCSFGLHARGVELMRFTTFFGPDSPYPA 240
Db	181	LPPRARSLSKGFV	TSVVAFPDTSKTQVQTDNSCSFGLHARGVELMRFTTFFGPDSPYPA 240
Qy	241	HARCOWALRGDAD	SVLSITFRSEDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYPPS 300
Db	241	HARCOWALRGDAD	SVLSITFRSEDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYPPS 300
Qy	301	YNLTFFHSSQNVLL	ITLTINTERRHPGFEATFFQPRMSSCGGRURKKAQGTENSPYPGHY 360
Db	301	YNLTFFHSSQNVLL	ITLTINTERRHPGFEATFFQPRMSSCGGRURKKAQGTENSPYPGHY 360
Qy	361	PPNIDCTWNIEVNN	QHVKYSKFFYYLLEPGVPAGTCKPDYVEINGEKYCSERSQFVVTS 420
Db	361	PPNIDCTWNIEVNN	QHVKYSKFFYYLLEPGVPAGTCKPDYVEINGEKYCSERSQFVVTS 420
Qy	421	NSNKITVRPHSDQ	SYDTGTGLAEVLSYDSDPCPGQTCRTGRCIRKELRCDGWADCTDH 480
Db	421	NSNKITVRPHSDQ	SYDTGTGLAEVLSYDSDPCPGQTCRTGRCIRKELRCDGWADCTDH 480
Qy	481	SDELNCSCDAGHQT	CKNKECKPLFWYCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK 540
Db	481	SDELNCSCDAGHQT	CKNKECKPLFWYCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK 540
Qy	541	SOQNGKDDCGSDS	DEASCPCNVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDSEK 600
Db	541	SOQNGKDDCGSDS	DEASCPCNVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDSEK 600
Qy	601	DCDGLSFTRQARV	VGGTTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYID 660
Db	601	DCDGLSFTRQARV	VGGTTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYID 660
Qy	661	DRGFYSDPTQWTA	FLGLHDQSORSAFVQERRLKRITISHPFFNFDFDYDIALLELEKP 720
Db	661	DRGFYSDPTQWTA	FLGLHDQSORSAFVQERRLKRITISHPFFNFDFDYDIALLELEKP 720
Qy	721	AEYSSWVRPCLDP	ASHVFPAGKAIWVTGHTGYGTGALILQKGBIRVINTTCENLL 780
Db	721	AEYSSWVRPCLDP	ASHVFPAGKAIWVTGHTGYGTGALILQKGBIRVINTTCENLL 780
Qy	781	PQOITPRMCMVGL	SGVDSCQSDGGLSSVEADGRIFOAGVYVSWGCGAQRNKPQVYT 840
Db	781	PQOITPRMCMVGL	SGVDSCQSDGGLSSVEADGRIFOAGVYVSWGCGAQRNKPQVYT 840
Qy	841	RLPLFRDWIKENT	GV 855
Db	841	RLPLFRDWIKENT	GV 855

RESULT 8
ABP56619
ID ABP56619 standard; protein; 855 AA.
XX
AC ABP56619;
XX

RESULT 8
 ABP56619
 ID ABP56619 standard; protein; 855 AA.
 XX
 AC ABP56619;
 XX

DT 24-MAR-2003 (first entry)
 XX Human membrane-type serine protease MTSpl1 protein SEQ ID NO:2.
 DE Human; membrane-type serine protease; enzyme; MTSpl10; cytosolic;
 KW type-II membrane-type serine protease; neoplastic disease; tumour; MTSpl1;
 KW matriptase.
 XX Homo sapiens.
 OS WO200292841-A2.
 PN 21-NOV-2002.
 PD 14-MAY-2002; 2002WO-US015332.
 PF 14-MAY-2001; 2001US-0291001P.
 PR (CORV-) CORVAS INT INC.
 XX Madison EL, Yeh J;
 FI WPI; 2003-129309/12.
 DR N-PSDB; ABZ22450.
 XX
 PT New polypeptides comprising the protease domain of a type-II membrane-
 type serine protease (MTCPl10), or its mutants, useful for diagnosing
 PT neoplasms or malignancies, or for screening for MTCPl10 inhibitors for
 PT treating such diseases.
 XX
 PS Disclosure; Page 181-183; 198pp; English.
 XX
 CC The present invention describes a polypeptide comprising a purified
 CC single or two chain polypeptide, which comprises the protease domain of a
 CC type-II membrane-type serine protease (MTCPl10) or its catalytically
 CC active portion, or a mutant of it, where up to 50 % of the amino acids
 CC are replaced with another amino acid, and the resulting polypeptide is a
 CC single chain or two chain polypeptide that has a catalytic activity of at
 CC least 1-10 % of the unmutated polypeptide. MTCPl10 has cytosolic
 CC activity. The polypeptide containing the protease domain of the MTCPl10 is
 CC useful for detecting a neoplastic disease, and for diagnosing the
 CC presence of a pre-malignant lesion, a malignancy, or other pathological
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
 CC effectiveness. An inhibitor of the polypeptide containing the protease
 CC domain of MTCPl10 is useful for treating or preventing neoplastic disease
 CC in a mammal. An inhibitor of the activation cleavage of the zymogen form
 CC of the MTCPl10 polypeptide is useful for inhibiting tumour initiation,
 CC growth or progression, or treating (pre-)malignant conditions of the e.g.
 CC breast, cervix, prostate, lung, ovary or colon. The present sequence
 CC represents human MTSpl1 (also known as matriptase), which is used in an
 CC example from the present invention
 XX
 SQ Sequence 855 AA;

Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEGEVFLPVNNVKYKEHGPGRWVLA 60
 Db 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEGEVFLPVNNVKYKEHGPGRWVLA 60

Qy 61 VLIGLLVLLVIGLGVVHQLQVRDVRQKVGNGYMRITNENFVDAYENSNSTFVSLASKV 120
 Db 61 VLIGLLVLLVIGLGVVHQLQVRDVRQKVGNGYMRITNENFVDAYENSNSTFVSLASKV 120

Qy 121 KDALKLLYSYGVPLFGPYHKSASVATFSEGSVIAYTWSEFSIPQHLVEAEERVMAEERVVM 180
 Db 121 KDALKLLYSYGVPLFGPYHKSASVATFSEGSVIAYTWSEFSIPQHLVEAEERVMAEERVVM 180

Qy 181 LPPRARSLSKSFVTSVVAFFPTDSKTQVORTDNQSCSFGHARGVELMRFTTGPFPDPSYPA 240
 Db 181 LPPRARSLSKSFVTSVVAFFPTDSKTQVORTDNQSCSFGHARGVELMRFTTGPFPDPSYPA 240

Db 181 LPPRARSLSKSFVTSVVAFFPTDSKTQVORTDNQSCSFGHARGVELMRFTTGPFPDPSYPA 240
 Qy 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPPS 300
 Db 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPPS 300
 Qy 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQPLRMSSCGGRLRKAQGTNFSPPYPGHY 360
 Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQPLRMSSCGGRLRKAQGTNFSPPYPGHY 360
 Qy 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
 Db 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
 Qy 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
 Db 421 NSNKITVRFHSDOSYTDGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
 Qy 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNSEDGQSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDSDVNDGDNSEDGQSCPAQTFRCNSGKCLSK 540
 Qy 541 SQQNGKDCDGDSDSEASCPKVNVTCTKTYRCLNGCLCLSKGNPECDGKDCSDGSEK 600
 Db 541 SQQNGKDCDGDSDSEASCPKVNVTCTKTYRCLNGCLCLSKGNPECDGKDCSDGSEK 600
 Qy 601 DDCGLRSFTRQARVVGTDADSGEWPQVSLHALGQGHICGASLISPNMLVSAHCVYD 660
 Db 601 DDCGLRSFTRQARVVGTDADSGEWPQVSLHALGQGHICGASLISPNMLVSAHCVYD 660
 Qy 661 DRGFYSDPTQWTAFLGLHDQSORSPAGVQERLKRITISHPFFNDFDIDALLELEK 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSORSPAGVQERLKRITISHPFFNDFDIDALLELEK 720
 Qy 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGBIRVNTQTCENLL 780
 Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGBIRVNTQTCENLL 780
 Qy 781 PQQITPRMCMCVGLSGVDSCQDSCGSLSSVEADGRIFOAGVVSWDGCAQRNKPQVYT 840
 Db 781 PQQITPRMCMCVGLSGVDSCQDSCGSLSSVEADGRIFOAGVVSWDGCAQRNKPQVYT 840

RESULT 9
 AAO30146
 ID AAO30146 standard; protein; 855 AA.
 XX AAO30146;
 AC AAO30146;
 XX
 DT 03-SEP-2003 (first entry)
 XX Human membrane-type serine protease MTSpl1 protein.
 DE Serine protease 17; CUSP17; tumour; cancer; antisense therapy; prostate;
 KW breast; cervix; lung; ovary; colon; Gene therapy; human; enzyme; MTSpl1;
 KW membrane-type serine protease; matriptase.
 XX Homo sapiens.
 OS WO2003044179-A2.
 PN 30-MAY-2003.
 PD 20-NOV-2002; 2002WO-US037626.
 PF 20-NOV-2001; 2001US-0332015P.
 XX (CORV-) CORVAS INT INC.
 XX

PI Madison EL, Ong EO;
 XX WPI; 2003-449816/42.
 DR N-PSDB; AAL60792.
 XX
 PT New substantially purified serine protease 17 polypeptide and encoding
 PT nucleic acid, useful for diagnosing and treating tumor conditions and/or
 PT cancer, particularly of the breast, cervix, prostate, lung, ovary or
 PT colon.
 XX
 XX
 PS Disclosure; Page 175-177; 189pp; English.
 XX
 CC The invention relates to serine protease 17 polypeptide designated CVSP17
 CC and its corresponding nucleic acid sequence. The invention also relates
 CC to a method using CVSP17 protein to identify compounds that modulate its
 CC protease activity. The method is useful for preventing, diagnosing and
 CC treating disorders related to the serine protease 17 activity, such as
 CC tumour conditions and/or cancer, particularly of the breast, prostate,
 CC cervix, lung, ovary or colon. CVSP17 DNA is used in gene therapy and in
 CC antisense therapy. The present sequence is human membrane-type serine
 CC protease MTSP1 (also called matrilysin) protein. This sequence is used to
 CC illustrate the method of the invention
 XX
 XX Sequence 855 AA;
 SQ
 Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGKDPFGAGLKYNRHEKVGLEGEVEFLPVNNVKKVKKHGFGRWVLA 60
 DB 1 MGSDRARKGGGKDPFGAGLKYNRHEKVGLEGEVEFLPVNNVKKVKKHGFGRWVLA 60
 QY 61 VLIGLLVLLGIGFLVHQLYRDVQVNGYMTITNENFVDAYENSNTFEVSLASKV 120
 DB 61 VLIGLLVLLGIGFLVHQLYRDVQVNGYMTITNENFVDAYENSNTFEVSLASKV 120
 QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYYWFSPISQHLVEAEAEVMAEERYVM 180
 DB 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIAYYWFSPISQHLVEAEAEVMAEERYVM 180
 QY 181 LPPRARSLSKSVFTSVAPFPTDSKTVQRTQDNCSFGLHARGVELMRTTFGPDSPYPA 240
 DB 181 LPPRARSLSKSVFTSVAPFPTDSKTVQRTQDNCSFGLHARGVELMRTTFGPDSPYPA 240
 QY 241 HARCOWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSMPEHALVOLCTYPPS 300
 DB 241 HARCOWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSMPEHALVOLCTYPPS 300
 QY 301 YNLTFHSSQNVLLITLTNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFTNSPYPGHY 360
 DB 301 YNLTFHSSQNVLLITLTNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFTNSPYPGHY 360
 QY 361 PPNIDCTWNIENPVNQHVKVSEKPYLLPGLPGVAGTCKDYVEINGEKYCGERSQFVVT 420
 DB 361 PPNIDCTWNIENPVNQHVKVSEKPYLLPGLPGVAGTCKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKTIVRPHSDQSYTDGFLAEYLSYSDSPCPGQFTCRGTGRCKRLCDGWADCTDH 480
 DB 421 NSNKTIVRPHSDQSYTDGFLAEYLSYSDSPCPGQFTCRGTGRCKRLCDGWADCTDH 480
 QY 481 SDELNCSADAGHQTCKNFKPLFWCDVNDGDNDEGSCSPAQTFRCSNGKCLSK 540
 DB 481 SDELNCSADAGHQTCKNFKPLFWCDVNDGDNDEGSCSPAQTFRCSNGKCLSK 540
 QY 541 SQQCNKDDCGDSDSEACPKVNVCTKTYRCLNGLCLSKGNPCDKCKDCSDGSDK 600
 DB 541 SQQCNKDDCGDSDSEACPKVNVCTKTYRCLNGLCLSKGNPCDKCKDCSDGSDK 600
 QY 601 CDCGLRSFTROARVVGGTDADEGEMWQVSLHALGQGHICGASLISPNLVSAAHCYID 660
 DB 601 CDCGLRSFTROARVVGGTDADEGEMWQVSLHALGQGHICGASLISPNLVSAAHCYID 660

QY 661 DRGFYSDDPTQWTAFLGLHDQSRSAFGVQERRLRIISHFFNFDFDXDIALLELEKP 720
 DB 661 DRGFYSDDPTQWTAFLGLHDQSRSAFGVQERRLRIISHFFNFDFDXDIALLELEKP 720
 QY 721 AEYSSMRVPCICLPDASHVFPAGKAIWVTGHWGTYGGTGALILQKGEIRVINQTTCE 780
 DB 721 AEYSSMRVPCICLPDASHVFPAGKAIWVTGHWGTYGGTGALILQKGEIRVINQTTCE 780
 QY 781 PQQITPRMCMVGLSGVDSGQDSGGPLSSVEADGRIFOAGVVSWSGDCQORNKPGVYT 840
 DB 781 PQQITPRMCMVGLSGVDSGQDSGGPLSSVEADGRIFOAGVVSWSGDCQORNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855
 RESULT 10
 ID AAE29820 standard; protein; 855 AA.
 XX AAE29820;
 XX 24-FEB-2003 (first entry)
 XX Human membrane-type serine protease 1 (MTSP1).
 XX Human; type II membrane-type serine protease 9; tumour; transgenic;
 XX type II transmembrane serine protease; enzyme; gene therapy; MTSP9;
 XX neoplastic disease; transgenic animal; membrane-type serine protease 1;
 XX TSP; MTSP1; matrilysin.
 XX Homo sapiens.
 XX WO20027267-A2.
 XX 03-OCT-2002.
 XX 27-MAR-2002; 2002WO-US009611.
 XX 27-MAR-2001; 2001US-0279228P.
 XX 15-MAY-2001; 2001US-0291501P.
 XX (CORV-) CORVAS INT INC.
 XX Madison EL, Ong EO;
 XX WPI; 2003-018940/01.
 XX N-PSDB; AAD47225.
 XX New substantially purified single or two-chain type II membrane-type
 XX serine protease 9 (MTSP9) polypeptide, useful for monitoring tumor
 XX progression, inhibiting tumor initiation, or treating a malignant or pre-
 XX malignant condition.
 XX Disclosure; Page 183-185; 199pp; English.
 XX The invention relates to type II membrane-type serine protease 9 (MTSP9)
 XX polypeptides and polynucleotides. MTSP9 belongs to type II transmembrane
 XX serine protease (TSP) family. Sequences of the invention and their
 XX antibodies are useful for diagnosing, treating or preventing neoplastic
 XX disease in mammals. They are useful for monitoring tumor progression,
 XX inhibiting tumor initiation, growth or progression or treating malignant
 XX or pre-malignant conditions. Transgenic animals of the invention are
 XX useful in animal models of tumour initiation, growth and/or progression
 XX models. The invention is also useful in gene therapy. The present
 XX sequence is human membrane-type serine protease 1 (MTSP1). MTSP1 also
 XX referred as matrilysin is a member of the TSP family
 XX Sequence 855 AA;
 SQ
 Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy 1	MGSDRARKGGGKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKVKEHGPGRWVLAA 60
Db 1	MGSDRARKGGGKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKVKEHGPGRWVLAA 60
Qy 61	VLIGLLLVLLGIGFLVHMLQYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Db 61	VLIGLLLVLLGIGFLVHMLQYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Qy 121	KDALKLLYSGVPLGPHYKESAVTAFSEGSLVAYWSEFSIPOHLVEEAERVAERVVM 180
Db 121	KDALKLLYSGVPLGPHYKESAVTAFSEGSLVAYWSEFSIPOHLVEEAERVAERVVM 180
Qy 181	LPFRARSLKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTFPGFDPSPYPA 240
Db 181	LPFRARSLKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTFPGFDPSPYPA 240
Qy 241	HARQWALRGDADSVLSLTFRSFDLASCDSRGSIDLVTYNTLSPMEPHALVOLCGTYPYS 300
Db 241	HARQWALRGDADSVLSLTFRSFDLASCDSRGSIDLVTYNTLSPMEPHALVOLCGTYPYS 300
Qy 301	YNLTFFHSSQNVLLITLTNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Db 301	YNLTFFHSSQNVLLITLTNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Qy 361	PPNIDCTWNIENVNQHVKVSFKFFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db 361	PPNIDCTWNIENVNQHVKVSFKFFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy 421	NSNKITVRPHSDQSYTDGFLAEYLSYDSSDPCPGQFTCRGRCIRKELCDGWADCTDH 480
Db 421	NSNKITVRPHSDQSYTDGFLAEYLSYDSSDPCPGQFTCRGRCIRKELCDGWADCTDH 480
Qy 481	SDELNCSDAGHGFCKKNKCKPLFWCDVNDGDNDEGSCPAQTPRCSNGKCLSK 540
Db 481	SDELNCSDAGHGFCKKNKCKPLFWCDVNDGDNDEGSCPAQTPRCSNGKCLSK 540
Qy 541	SQOCNGKDDCGSDSEASCPKVVVVTCTKHYRCLNGLCLSKNPECDGKEDCDGSD 600
Db 541	SQOCNGKDDCGSDSEASCPKVVVVTCTKHYRCLNGLCLSKNPECDGKEDCDGSD 600
Qy 601	DCDCGLRFTROARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601	DCDCGLRFTROARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661	DRGFRYSPTQWTAFLGLHDQSORSAPGVQERRLKRIISHPPFNDFTFDYDIALLEK 720
Db 661	DRGFRYSPTQWTAFLGLHDQSORSAPGVQERRLKRIISHPPFNDFTFDYDIALLEK 720
Qy 721	AEYSSMVRPICLEADASHVFPAGKAIWVTGWGHTQYGGTGALILQGEIRVINQTCENLL 780
Db 721	AEYSSMVRPICLEADASHVFPAGKAIWVTGWGHTQYGGTGALILQGEIRVINQTCENLL 780
Qy 781	PQOITPRMVCYGLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCQARNKPGYTT 840
Db 781	PQOITPRMVCYGLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCQARNKPGYTT 840
Qy 841	RLPLFRDWIKENTGV 855
Db 841	RLPLFRDWIKENTGV 855

RESULT 11

ID	AAE29791	standard; protein; 855 AA.
XX	AAE29791;	
AC	AAE29791;	
XX	XX	
XX	XX	
DT	24-FEB-2003	(first entry)
XX	XX	Human membrane-type serine protease, MTSP1.
DE	DE	
XX	XX	

KW	Human; serine protease 14; CVSP14; cancer; malignancy; breast; colon; gene therapy; cytosstatic; membrane-type serine protease; enzyme; MTSP1.
KW	Homo sapiens.
OS	WO200277263-A2.
FN	03-OCT-2002.
PD	20-MAR-2002; 2002WO-US009039.
XX	22-MAR-2001; 2001US-0278166P.
XX	(CORV-) CORVAS INT INC.
PA	Madison EL, Yeh J;
XX	WPI; 2003-018938/01.
PI	N-PSDB; AAD47180.
XX	New purified CVSP14 polypeptide and encoding nucleic acid molecule, useful for diagnosing, preventing and/or treating disorders, such as cancers and malignancies of the breast, cervix, prostate, lung, ovary or colon.
DR	Disclosure; Page 171-173; 185pp; English.
XX	The invention relates to transmembrane serine protease 14 (CVSP14), its nucleic acid sequence and the method based on them. The methods and compositions of the invention are useful for diagnosing, preventing or and/or treating conditions associated with the aberrant expression or activity of the CVSP14 polypeptide, such as cancers and malignancies of the breast, cervix, prostate, lung, ovary or colon. The methods are also useful for identifying compounds that will modulate the protease activity of CVSP14 polypeptide and monitoring tumour progression and/or therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present sequence is human membrane-type serine protease, MTSP1
CC	Sequence 855 AA;
XX	
SS	Query Match 100.0%; Score 4681; DB 6; Length 855;
XX	Best Local Similarity 100.0%; Pred. No. 2.1e-307; Indels 0; Gaps 0;
CC	Matches 855; Conservative 0; Mismatches 0;

Qy 1	MGSDRARKGGGKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKVKEHGPGRWVLAA 60
Db 1	MGSDRARKGGGKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKVKEHGPGRWVLAA 60
Qy 61	VLIGLLLVLLGIGFLVHMLQYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Db 61	VLIGLLLVLLGIGFLVHMLQYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Qy 121	KDALKLLYSGVPLGPHYKESAVTAFSEGSLVAYWSEFSIPOHLVEEAERVAERVVM 180
Db 121	KDALKLLYSGVPLGPHYKESAVTAFSEGSLVAYWSEFSIPOHLVEEAERVAERVVM 180
Qy 181	LPFRARSLKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTFPGFDPSPYPA 240
Db 181	LPFRARSLKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTFPGFDPSPYPA 240
Qy 241	HARQWALRGDADSVLSLTFRSFDLASCDSRGSIDLVTYNTLSPMEPHALVOLCGTYPYS 300
Db 241	HARQWALRGDADSVLSLTFRSFDLASCDSRGSIDLVTYNTLSPMEPHALVOLCGTYPYS 300
Qy 301	YNLTFFHSSQNVLLITLTNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Db 301	YNLTFFHSSQNVLLITLTNTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Qy 361	PPNIDCTWNIENVNQHVKVSFKFFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db 361	PPNIDCTWNIENVNQHVKVSFKFFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy 421	NSNKITVRPHSDQSYTDGFLAEYLSYDSSDPCPGQFTCRGRCIRKELCDGWADCTDH 480

Db 421 NSNKITVRHSDQSYDTGFLAEYLSYSSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Qy 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 Qy 541 SQOCNGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDEK 600
 Db 541 SQOCNGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDEK 600
 Qy 601 DCDGLRSFTQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Qy 661 DRGFRYSDDPTQWTAFLGLHDQSQRSAPGVQERLRKRIISHPPFNDFTDYDIALLELEKP 720
 Db 661 DRGFRYSDDPTQWTAFLGLHDQSQRSAPGVQERLRKRIISHPPFNDFTDYDIALLELEKP 720
 Qy 721 AYSWMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Db 721 AYSWMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Qy 781 PQOITPRMCMCVGLSGGVDSGCGPLSSVEADGRIFOAGVSWGDCQQRNKPQYVT 840
 Db 781 PQOITPRMCMCVGLSGGVDSGCGPLSSVEADGRIFOAGVSWGDCQQRNKPQYVT 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 12

ABP72376
 ID ABP72376 standard; protein; 855 AA.
 XX AC ABP72376;
 XX DT 13-MAY-2003 (first entry)
 XX DE Transmembrane serine protease 1 (MTP1).
 KW Human; transmembrane serine protease 1; MTP1; MTP20; enzyme;
 KW cystostatic; dermatological; cardiant; vulnary; ophthalmological;
 KW gene therapy.
 XX OS Homo sapiens.
 XX PN WO2003004681-A2.
 XX PD 16-JAN-2003.
 XX PF 03-JUL-2002; 2002WO-US021208.
 XX PR 03-JUL-2001; 2001US-0302939P.
 XX PA (CORV-) CORVAS INT INC.
 XX PI Madison EL, Ong EO;
 XX DR WPI; 2003-239207/23.
 XX DR N-PSDB; ABZ58500.
 XX New type-II membrane-type serine protease 20 polypeptides, useful for
 PT preparing a medicament for diagnosing, treating or preventing cancer,
 PT dermatological disorders, aberrant wound repairs or crest syndromes.
 XX PS Disclosure; Page 198-200; 216pp; English.
 XX CC The present sequence is the protein sequence of human type II
 CC transmembrane serine protease 1 (MTP1). The invention relates to novel
 CC human type II transmembrane serine protein 20 (MTP20) (see ABP72374) and
 CC nucleic acids encoding it (see ABZ58499). Also claimed are methods of

CC inhibiting tumour initiation, growth or progression by inhibiting MTP20
 CC activity, and of treating or preventing a disease or disorder associated
 CC with undesired and/or uncontrolled angiogenesis or neovascularisation,
 CC especially undesired angiogenesis associated with solid neoplasms,
 CC vascular malformations and cardiovascular disorders, chronic inflammatory
 CC diseases, aberrant wound repairs, circulatory disorders, crest syndromes,
 CC dermatological disorders and ocular disorders using an inhibitor of MTP20
 XX
 SQ Sequence 855 AA;
 Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307; Indels 0; Gaps 0;
 Matches 855; Conservative 0; Mismatches 0;
 Qy 1 MGSDBAKGGGGPDXFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKYKKGPGRWVLA 60
 Db 1 MGSDBAKGGGGPDXFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKYKKGPGRWVLA 60
 Qy 61 VLIGLLVLVIGLGVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 Db 61 VLIGLLVLVIGLGVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 Qy 121 KDALKLYSGVPFLGYPYHKSATVAFSEGSVIAIYWSSEFSIPOHLVEAEVMAEERVVM 180
 Db 121 KDALKLYSGVPFLGYPYHKSATVAFSEGSVIAIYWSSEFSIPOHLVEAEVMAEERVVM 180
 Qy 181 LPPRARSLSKSPVTSVVAFFTDSTKVQRTQDNSCSFGLHARGVELMRTFTPGFDPSPY 240
 Db 181 LPPRARSLSKSPVTSVVAFFTDSTKVQRTQDNSCSFGLHARGVELMRTFTPGFDPSPY 240
 Qy 241 HARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSPMEHALVOLCGTYP 300
 Db 241 HARCQWALRGDADSVLSITFRSFDLASCDERGSDLVTYNTLSPMEHALVOLCGTYP 300
 Qy 301 YNLTFFHSSQNVLLITLTINTERRHPGFEATFFQPLRMSSCGRLRKAQGTNSPYPGHY 360
 Db 301 YNLTFFHSSQNVLLITLTINTERRHPGFEATFFQPLRMSSCGRLRKAQGTNSPYPGHY 360
 Qy 361 PPNIDCTWNIIEVPNNQHVKSFKFFYLLRPGVPAGTCKDYVEINGEKYQGERSO 420
 Db 361 PPNIDCTWNIIEVPNNQHVKSFKFFYLLRPGVPAGTCKDYVEINGEKYQGERSO 420
 Qy 421 NSNKITVRHSDQSYDTGFLAEYLSYSSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Db 421 NSNKITVRHSDQSYDTGFLAEYLSYSSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Qy 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 Qy 541 SQOCNGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDEK 600
 Db 541 SQOCNGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDEK 600
 Qy 601 DCDGLRSFTQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Qy 661 DRGFRYSDDPTQWTAFLGLHDQSQRSAPGVQERLRKRIISHPPFNDFTDYDIALLELEKP 720
 Db 661 DRGFRYSDDPTQWTAFLGLHDQSQRSAPGVQERLRKRIISHPPFNDFTDYDIALLELEKP 720
 Qy 721 AYSWMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Db 721 AYSWMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Qy 781 PQOITPRMCMCVGLSGGVDSGCGPLSSVEADGRIFOAGVSWGDCQQRNKPQYVT 840
 Db 781 PQOITPRMCMCVGLSGGVDSGCGPLSSVEADGRIFOAGVSWGDCQQRNKPQYVT 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 13	
AD97551	
ID	AD97551 standard; protein; 855 AA.
XX	
AC	AD97551;
XX	
DT	04-DEC-2003 (first entry)
XX	
DE	Human MTSPL, SEQ ID NO:2.
XX	
KW	Human; type I transmembrane serine protease 1; MTSPL; matrilysin; serine protease; protease cleavage activation; diagnostic marker; neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary; cytostatic; gene therapy; drug screening; tumour progression; monitoring; enzyme.
KW	
KW	
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO2003031585-A2.
XX	
PD	17-APR-2003.
XX	
PF	08-OCT-2002; 2002WO-US032417.
XX	
PR	09-OCT-2001; 2001US-0328530P.
XX	
PA	(CORV-) CORVAS INT INC.
XX	
PI	Madison EL, Yeh J;
XX	
DR	WPI; 2003-393442/37.
XX	
DR	N-PSDB; AD97550.
XX	
PT	New purified single- or two-chain polypeptide, useful for diagnosing, preventing or treating cancer (e.g. colon cancer), comprises a protease domain of a type-I membrane-type serine protease 25 or its catalytically active portion.
PT	
XX	
PS	Disclousure; Page 179-181; 97pp; English.
XX	
CC	The invention relates to human type I transmembrane serine protease 25 (MTSP25; AD97551) and polypeptides derived from it (e.g., AD97555). The MTSPL gene is located on chromosome 12. MTSPL is a serine protease which is activated by protease cleavage to yield a two-chain protease comprising an A chain and a B chain linked by a disulphide bond. MTSPL is expressed or is active in tumour cells, and can therefore be used as a diagnostic marker for certain cancers. The invention also encompasses nucleic acids encoding an MTSPL polypeptide (AD97554, AD97571, AD97554); nucleic acid vectors and host cells comprising an MTSPL polynucleotide; a MTSPL knockout animal; and an antibody specific for MTSPL. MTSPL polypeptides are useful in diagnosing, preventing or treating neoplastic diseases, such as cancer of the breast, cervix, prostate, lung, ovary or colon. MTSPL polypeptides may also be used for identifying compounds that modulate the protease activity of the polypeptide and for monitoring tumour progression and/or therapeutic effectiveness. The present sequence represents the related protein, MTSPL (also known as matrilysin).
XX	
SQ	Sequence 855 AA;
Query Match	100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity	100.0%; Pred. No. 2.1e-307;
Matches 855; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MGSDRARKGKGGKDFGAGLKYNSRHEKVNGLSEGVFLPNNVKYKHGPGRWVLA 60
Db	1 MGSDRARKGKGGKDFGAGLKYNSRHEKVNGLSEGVFLPNNVKYKHGPGRWVLA 60
Qy	61 VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNTFVSLASKV 120

Db	61	VLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNTFVSLASKV	120
Qy	121	KDALKLYSGVPLPGPYHKSASVATFSEGSAVIAIYTWSEFSIQHVLVEAEVMAERVVM	180
Db	121	KDALKLYSGVPLPGPYHKSASVATFSEGSAVIAIYTWSEFSIQHVLVEAEVMAERVVM	180
Qy	181	LPPRARSLSKGFVVTSSVAFPTDSKTQVTDQNSCSFGLHARGVELMRFTEGFPDPSYPA	240
Db	181	LPPRARSLSKGFVVTSSVAFPTDSKTQVTDQNSCSFGLHARGVELMRFTEGFPDPSYPA	240
Qy	241	HARCOWALRGDADSVLSITFRSPDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYP	300
Db	241	HARCOWALRGDADSVLSITFRSPDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYP	300
Qy	301	YNLTFFHSSQNVLLITLTNTERRHPGEATFFQPRMSSCGGRURKKAQGTNSPYPGHY	360
Db	301	YNLTFFHSSQNVLLITLTNTERRHPGEATFFQPRMSSCGGRURKKAQGTNSPYPGHY	360
Qy	361	PPNIDCTWNIIEVNNQHVSKFYYLLEPGVPAGTCTGRCIRKELRCGWDADCTDH	420
Db	361	PPNIDCTWNIIEVNNQHVSKFYYLLEPGVPAGTCTGRCIRKELRCGWDADCTDH	420
Qy	421	NSNKITVRPHSDQSYTDTGFLAEVLSYDSSDPGPGTCTGRCIRKELRCGWDADCTDH	480
Db	421	NSNKITVRPHSDQSYTDTGFLAEVLSYDSSDPGPGTCTGRCIRKELRCGWDADCTDH	480
Qy	481	SDELNCSCDAGHQTCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRCNSGKCLSK	540
Db	481	SDELNCSCDAGHQTCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRCNSGKCLSK	540
Qy	541	SOQNGKDCDGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK	600
Db	541	SOQNGKDCDGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK	600
Qy	601	DCDGLSFRQARVVGSTDADEGEWPQVSLHALGQGHICGASLISPNVLVSAHCYID	660
Db	601	DCDGLSFRQARVVGSTDADEGEWPQVSLHALGQGHICGASLISPNVLVSAHCYID	660
Qy	661	DRGFRYSDPTQWTAFLGLHDQSORAPGVORERLKRILSHPPFNDFPDYDIALLELEKP	720
Db	661	DRGFRYSDPTQWTAFLGLHDQSORAPGVORERLKRILSHPPFNDFPDYDIALLELEKP	720
Qy	721	AEYSSMVRPCLPDASHVFPAGKAIWVTGHTGYGCTGALILQKGBIRVNTQTCENLL	780
Db	721	AEYSSMVRPCLPDASHVFPAGKAIWVTGHTGYGCTGALILQKGBIRVNTQTCENLL	780
Qy	781	POQITPRMCMVGLSGVSDSCQDSCGGLSSVADGRIFQAGVSWDGCQRNKPQVYT	840
Db	781	POQITPRMCMVGLSGVSDSCQDSCGGLSSVADGRIFQAGVSWDGCQRNKPQVYT	840
Qy	841	RLPLFRDWIKENTGV 855	
Db	841	RLPLFRDWIKENTGV 855	
XX		RESULT 14	
AD97551		AD97551	
ID	AD97551 standard; protein; 855 AA.		
XX		AC	AD97551;
XX		DT	22-APR-2004 (first entry)
XX		DE	Human cell surface protease #1.
XX		KW	therapeutic agent; plasmin; protease specific antigen; PSA; cell-surface protease-associated disease; cancer; ocular disease; cardiovascular disease; chronic inflammatory disease; wound; circulatory disorder; dermatological disorder; rheumatoid arthritis; psoriasis; diabetic retinopathy; pterygium; excimer laser surgery scarring; glaucoma filtering surgery scarring; macular degeneration; crest syndrome; solid neoplasm; vascular tumour; melanoma; Kaposi's sarcoma; human; cell surface protease.

xx	This sequence represents a human type II membrane-type serine protease 1 (MTSP1). The MTSP1 protein sequence was used in the isolation of the coding sequence of MTSP12, which has been isolated to chromosome 19. MTSP12 does not include the sequence of amino acids Arg-Lys-His-Leu-Pro-Arg-Pro-Ala ADG65347. The MTSP12 coding sequence was identified by using the protein sequence of the protease domain of MTSP1 to search the human HTGS database which produced three serine proteases, MTSP12-PD1, MTSP12-PD2 and MTSP12-PD3. EST's were identified which matched portions of MTSP12-PD1, MTSP12-PD2 and MTSP12-PD3. MTSP12 polypeptides, proteins and nucleic acids are useful in diagnostics, particularly for diagnosing lung, oesophageal, prostate, colon, ovary, cervix, breast and pancreas cancers. These are useful in immunoassays to detect, prognosis, diagnose, or monitor various conditions, diseases, and disorders affecting MTSP12 polypeptide expression, or monitor their treatment. Modulators of MTSP12 are useful for treating cancer, tumour and other neoplastic conditions.									
xx	SQ Sequence 855 AA;									
	Query Match	100.0%;	Score 4681;	DB 8;	Length 855;					
	Best Local Similarity	100.0%;	Pred. No. 2.1e-307;							
	Matches 855;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Qy	1	MGSDRARKGGGKDFGAGLKNSRHEKNGLEEGVEFLPVNNVKKVKGHGPRWVLA	60							
Db	1	MGSDRARKGGGKDFGAGLKNSRHEKNGLEEGVEFLPVNNVKKVKGHGPRWVLA	60							
Qy	61	VLIGLLLVLLGLFLVHLYRDRVQVFNQYMTNENFVDAVENSNSTEFVSLASKV	120							
Db	61	VLIGLLLVLLGLFLVHLYRDRVQVFNQYMTNENFVDAVENSNSTEFVSLASKV	120							
Qy	121	KDALKLLYSGVPLGPGYHKESAVTAFSEGSVIAYYWSFSPQHLVVEAERVMAE	180							
Db	121	KDALKLLYSGVPLGPGYHKESAVTAFSEGSVIAYYWSFSPQHLVVEAERVMAE	180							
Qy	181	LPPRARSLKSFVTSVVAFPDTSKTQRTQDNCSFGLHARGVELMRFTTGFDP	240							
Db	181	LPPRARSLKSFVTSVVAFPDTSKTQRTQDNCSFGLHARGVELMRFTTGFDP	240							
Qy	241	HARQWALRGDASVLSLTFRSFDLASCDERGSDLVTYNTLSPEPHALVOLCG	300							
Db	241	HARQWALRGDASVLSLTFRSFDLASCDERGSDLVTYNTLSPEPHALVOLCG	300							
Qy	301	YNLTFHSSQVLLITLITNTERRHPGFATFFQLPRMSSCGRLKKAQGTNSPY	360							
Db	301	YNLTFHSSQVLLITLITNTERRHPGFATFFQLPRMSSCGRLKKAQGTNSPY	360							
Qy	361	PNIDCTWNIENPNQHVKSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQ	420							
Db	361	PNIDCTWNIENPNQHVKSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQ	420							
Qy	421	NSNKITVRFHSDQSYTDGFLAEVLSYSDSDPCQFTCTGRCIRKELCDGWAD	480							
Db	421	NSNKITVRFHSDQSYTDGFLAEVLSYSDSDPCQFTCTGRCIRKELCDGWAD	480							
Qy	481	SDELNCSADAGHQFTCKNFKCKPLFWVCDVNDQDNDEQSCPAQTFRCNSGK	540							
Db	481	SDELNCSADAGHQFTCKNFKCKPLFWVCDVNDQDNDEQSCPAQTFRCNSGK	540							
Qy	541	SQCNKGKDDCGSDSEASCPKVVVTCYKHYRCLNGCLSKGNPECDGKEDCS	600							
Db	541	SQCNKGKDDCGSDSEASCPKVVVTCYKHYRCLNGCLSKGNPECDGKEDCS	600							
Qy	601	DCDCGLRSFTQARVVGTTDADEGEWPNQVSLHALGQGHICGASLISPNWL	660							
Db	601	DCDCGLRSFTQARVVGTTDADEGEWPNQVSLHALGQGHICGASLISPNWL	660							
Qy	661	DRGFYSDPQTQWTAFLGLHDQSQRSAFCVQERRLKRIISHPFFNDFTFDY	720							
Db	661	DRGFYSDPQTQWTAFLGLHDQSQRSAFCVQERRLKRIISHPFFNDFTFDY	720							
Qy	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLQKGEIRVIN	780							

Search completed: September 23, 2005, 12:55:36
Job time : 88 secs

Db	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLQKGEIRVIN	780
Qy	781	PQOITPRMNCVGFSLSGGVDSQGDSCGGPLSSVEADGRIFQAGVVS	840
Db	781	PQOITPRMNCVGFSLSGGVDSQGDSCGGPLSSVEADGRIFQAGVVS	840
Qy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855